

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 18, 2005, 21:11:10 ; Search time 165 Seconds
(without alignments)
560.217 Million cell updates/sec

Title: US-10-035-300A-2

Perfect score: 1225
Sequence: 1 MATPHINAEAGDFADVILMP.....TTFNDMIKIALESVLLGDKE 239

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_15Dec04:*
1: Geneseqp1980s:*
2: Geneseqp1990s:*
3: Geneseqp2000s:*
4: Geneseqp2001s:*
5: Geneseqp2002s:*
6: Geneseqp2003as:*
7: Geneseqp2003bs:*
8: Geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1225	100.0	239	6	AAE37217 Escherich
2	1224	99.9	239	6	ABR44457 Escherich
3	1222	99.8	239	4	AAB67584 Amino aci
4	1222	99.8	239	6	AAE35397 Escherich
5	1222	99.8	239	8	ADL27842 E coli de
6	1221	99.7	239	6	ABR44456 Escherich
7	1221	99.7	239	6	AAE35404 Escherich
8	1220	99.6	239	6	AAE35402 Escherich
9	1219	99.5	239	6	ABR44462 Escherich
10	1219	99.5	239	6	ABR44458 Escherich
11	1219	99.5	239	6	ABR44472 Escherich
12	1219	99.5	239	6	ABR44467 Escherich
13	1219	99.5	239	6	AAE35405 Escherich
14	1219	99.5	239	6	AAE35416 Escherich
15	1219	99.5	239	6	AAE35411 Escherich
16	1219	99.5	239	6	AAE35406 Escherich
17	1218	99.4	239	6	AAE37218 Escherich
18	1218	99.4	239	6	ABR44469 Escherich
19	1218	99.4	239	6	AAE35399 Escherich
20	1218	99.4	239	6	AAE35410 Escherich
21	1217	99.3	239	6	ABR44475 Escherich
22	1217	99.3	239	6	ABR44461 Escherich
23	1217	99.3	239	6	ABR44464 Escherich
24	1217	99.3	239	6	ABR44463 Escherich
25	1217	99.3	239	6	AAE35401 Escherich

26	1217	99.3	239	6	AAE35403	AAE35403 Escherich
27	1217	99.3	239	6	AAE35412	AAE35412 Escherich
28	1216	99.3	239	6	ABR44460	ABR44460 Escherich
29	1216	99.3	239	6	ABR44465	ABR44465 Escherich
30	1216	99.3	239	6	ABR44466	ABR44466 Escherich
31	1216	99.3	239	6	ABR44471	ABR44471 Escherich
32	1216	99.3	239	6	AAE35398	AAE35398 Escherich
33	1216	99.3	239	6	AAE35413	AAE35413 Escherich
34	1216	99.3	239	6	AAE35409	AAE35409 Escherich
35	1216	99.3	239	6	AAE35415	AAE35415 Escherich
36	1216	99.3	239	6	AAE35407	AAE35407 Escherich
37	1215	99.2	239	6	ABR44470	ABR44470 Escherich
38	1215	99.2	239	6	ABR44468	ABR44468 Escherich
39	1215	99.2	239	6	ABR44459	ABR44459 Escherich
40	1215	99.2	239	6	ABR44473	ABR44473 Escherich
41	1214	99.1	239	6	ABR44474	ABR44474 Escherich
42	1214	99.1	239	6	AAE35408	AAE35408 Escherich
43	1213	99.0	239	4	AAB67590	AAB67590 Amino aci
44	1211	98.9	239	6	AAE35414	AAE35414 Escherich
45	1162	94.9	243	7	ABO62285	ABO62285 Klebsiell

ALIGNMENTS

RESULT 1
AAE37217
ID AAE37217 standard; protein; 239 AA.
XX AAE37217;
AC AAE37217;
DT 07-AUG-2003 (first entry)
XX
DE Escherichia coli mutant PNP enzyme (M65V) #SEQ ID 2.
XX
KW Purine nucleoside phosphorylase; mutein; tumour; PNP; enzyme; cancer;
KW leukaemia; glioblastoma; lymphoma; melanoma; carcinoma; gene therapy.
XX
OS Escherichia coli.
OS Synthetic.
FH Key Location/Qualifiers
FT Misc-difference 65
FT /note= "this residue is Met in the wild-type sequence"
XX
XX WO2003035012-A2.
XX
PD 01-MAY-2003.
XX
XX 28-OCT-2002; 2002WO-US034626.
XX
XX 26-OCT-2001; 2001US-00035300.
XX
XX (UABR-) UAB RES FOUND.
XX PA (SOOR) SOUTHERN RES INST.
XX PA (CORR) CORNELL RES FOUND INC.
XX
XX PI Ealick SE, Parker WB, Seerist JA, Sorscher BJ;
XX DR WPI; 2003-421350/39.
XX DR N-PSDB; AAD56237.
XX
XX Novel purified mutant purine cleaving enzyme having different biological
XX activity than a wild-type purine cleaving enzyme, useful for treating a
XX pathological condition characterized by abnormal cell growth.
XX
XX Claim 9; Page 64-65; 67pp; English.
XX
XX The invention relates to mutant purine nucleoside phosphorylase (PNP)
XX enzymes and nucleic acid molecules encoding such enzymes. These mutant
XX enzymes have greater biological activity than wild-type enzymes. PNP
XX enzymes are useful for treating a pathological condition characterised by
XX abnormal cell growth such as that occurring in cancer of the skin,

CC breast, bone, colon, head or neck, kidney, larynx, lung, liver, ovary,
 CC nasopharynx, oesophagus, pancreas, prostate, bladder, rectum, stomach,
 CC thyroid, testicle and uterus as well as other conditions characterised by
 CC abnormal cell growth such as myeloid leukaemia, glioblastoma and
 CC lymphoma. Sequences of the invention are useful for treating metastatic
 CC solid tumours such as melanoma, pancreatic, liver or colonic carcinoma or
 CC to activate prodrugs in the treatment of cancer. They are useful for
 CC treating virally infected cells. They are also useful in gene therapy.
 CC The present sequence is Escherichia coli mutant PNP enzyme (M65V) amino
 CC acid sequence
 XX
 SQ Sequence 239 AA;
 Query Match 100.0%; Score 1225; DB 6; Length 239;
 Best Local Similarity 100.0%; Pred. No. 1.1e-123;
 Matches 239; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MATPHINAEMGDFADVLMFGDPLRAKYIAETFLDAREVNNVGMFGTGTGKRSIV 60
 DB 1 MATPHINAEMGDFADVLMFGDPLRAKYIAETFLDAREVNNVGMFGTGTGKRSIV 60
 QY 61 MGHGVGIPSCSIYTKELITDFGVKKIIRVSGCGAVLPHVKLRDVVIGGACTDSKNRIR 120
 DB 61 MGHGVGIPSCSIYTKELITDFGVKKIIRVSGCGAVLPHVKLRDVVIGGACTDSKNRIR 120
 QY 121 FKDHFAAIAADFDMVRNAVDAKALGIDARVGNLFSADLFYSPDGFVMEKYGILGVE 180
 DB 121 FKDHFAAIAADFDMVRNAVDAKALGIDARVGNLFSADLFYSPDGFVMEKYGILGVE 180
 QY 181 MEAGIYGVAEFGAKALITCTVSDHIRTHTHEQTAAERQTTFNMDIKIALESVLLGDKE 239
 DB 181 MEAGIYGVAEFGAKALITCTVSDHIRTHTHEQTAAERQTTFNMDIKIALESVLLGDKE 239
 RESULT 2
 ABR44457
 ID ABR44457 standard; protein; 239 AA.
 XX
 AC ABR44457;
 XX
 XX 07-AUG-2003 (first entry)
 DE Escherichia coli mutant PNP enzyme M65I.
 KW Purine nucleoside phosphorylase; mukein; tumour; PNP; enzyme; cancer;
 KW leukaemia; glioblastoma; lymphoma; melanoma; carcinoma; gene therapy.
 XX
 OS Escherichia coli.
 OS Synthetic.
 XX
 XX Key Location/Qualifiers
 FT Misc-difference 65 /note= "wild-type Met replaced with Ile"
 FT
 XX
 XX WO2003035012-A2.
 XX
 XX 01-MAY-2003.
 XX
 XX 28-OCT-2002; 2002WO-US034626.
 XX
 XX 26-OCT-2001; 2001US-00035300.
 XX
 XX (UABR-) UAB RES FOUND.
 PA (SOUR) SOUTHERN RES INST.
 PA (CORR) CORNELL RES FOUND INC.
 XX
 XX Balick SE, Parker WB, Seerist JA, Sorscher EJ;
 XX WPI; 2003-421350/39.
 XX
 XX Novel purified mutant purine cleaving enzyme having different biological
 PT activity than a wild-type purine cleaving enzyme, useful for treating a
 PT pathological condition characterized by abnormal cell growth.

XX Claim 13; Page; 67pp; English.
 PS
 XX The invention relates to mutant purine nucleoside phosphorylase (PNP)
 CC enzymes and nucleic acid molecules encoding such enzymes. These mutant
 CC enzymes have greater biological activity than wild-type enzymes. PNP
 CC enzymes are useful for treating a pathological condition characterised by
 CC abnormal cell growth such as that occurring in cancer of the skin,
 CC breast, bone, colon, head or neck, kidney, larynx, lung, liver, ovary,
 CC nasopharynx, oesophagus, pancreas, prostate, bladder, rectum, stomach,
 CC thyroid, testicle and uterus as well as other conditions characterised by
 CC abnormal cell growth such as myeloid leukaemia, glioblastoma and
 CC lymphoma. Sequences of the invention are useful for treating metastatic
 CC solid tumours such as melanoma, pancreatic, liver or colonic carcinoma or
 CC to activate prodrugs in the treatment of cancer. They are useful for
 CC treating virally infected cells. They are also useful in gene therapy.
 CC The present sequence is Escherichia coli mutant PNP enzyme (M65I) amino
 CC acid sequence. NOTE: This sequence is given in the specification, but may
 CC be created by making the wild-type sequence from the mutant sequence
 CC given in SEQ ID 2 (AAB37217) and then creating the desired mutation
 XX
 SQ Sequence 239 AA;
 Query Match 99.9%; Score 1224; DB 6; Length 239;
 Best Local Similarity 99.6%; Pred. No. 1.5e-123;
 Matches 239; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MATPHINAEMGDFADVLMFGDPLRAKYIAETFLDAREVNNVGMFGTGTGKRSIV 60
 DB 1 MATPHINAEMGDFADVLMFGDPLRAKYIAETFLDAREVNNVGMFGTGTGKRSIV 60
 QY 61 MGHGVGIPSCSIYTKELITDFGVKKIIRVSGCGAVLPHVKLRDVVIGGACTDSKNRIR 120
 DB 61 MGHGVGIPSCSIYTKELITDFGVKKIIRVSGCGAVLPHVKLRDVVIGGACTDSKNRIR 120
 QY 121 FKDHFAAIAADFDMVRNAVDAKALGIDARVGNLFSADLFYSPDGFVMEKYGILGVE 180
 DB 121 FKDHFAAIAADFDMVRNAVDAKALGIDARVGNLFSADLFYSPDGFVMEKYGILGVE 180
 QY 181 MEAGIYGVAEFGAKALITCTVSDHIRTHTHEQTAAERQTTFNMDIKIALESVLLGDKE 239
 DB 181 MEAGIYGVAEFGAKALITCTVSDHIRTHTHEQTAAERQTTFNMDIKIALESVLLGDKE 239
 RESULT 3
 AAB57584
 ID AAB57584 standard; protein; 239 AA.
 XX
 AC AAB57584;
 XX
 XX 29-MAY-2001 (first entry)
 DT
 XX
 DE Amino acid sequence of a purine nucleoside phosphorylase.
 XX
 KW Deoxyribonucleoside; deoxyribose 1-phosphate; thymidine phosphorylase;
 KW purine nucleoside phosphorylase; phosphopentose mutase;
 KW phosphopentose aldolase; fructose 1,6-diphosphate aldolase;
 KW deoxyribokinase; nucleoside 2-deoxyribosyltransferase.
 XX
 OS Escherichia coli.
 OS
 XX WO200114566-A2.
 XX
 XX 01-MAR-2001.
 XX
 XX 18-AUG-2000; 2000WO-EP008088.
 PF
 XX
 XX 20-AUG-1999; 99EP-00116425.
 XX
 XX (HOFF) ROCHE DIAGNOSTICS GMBH.
 PA (INSP) INST PASTEUR.
 PA (PHAR-) PHARMA-WALDHOF GMBH & CO KG.
 XX

PI Tischer W, Ihlenfeldt H, Barzu O, Sakamoto H, Pistotnik E;
PI Marliere P, Pochet S;
XX
DR N-PSDB; AAF55440.
DR WPI; 2001-235026/24.
XX

XX In vitro enzymatic synthesis of deoxyribonucleosides comprises reacting
PT deoxyribose 1-phosphate and a nucleobase to form a deoxyribonucleoside
PT and an inorganic phosphate.
XX

PS Disclosure; Page 45; 73pp; English.

XX The present sequence represents a purine nucleoside phosphorylase enzyme.
CC This enzyme is involved in the biosynthesis of deoxyribonucleosides, and
CC is used in the method of the invention. The specification describes a
CC method for the in vitro enzymatic synthesis of deoxyribonucleosides. The
CC method comprises reacting deoxyribose 1-phosphate and a nucleobase to
CC form a deoxyribonucleoside and an inorganic phosphate. Enzymes which may
CC be used in the method of the invention include thymidine phosphorylase,
CC purine nucleoside phosphorylase, phosphopentose mutase, phosphopentose
CC aldolase, fructose 1,6-diphosphate aldolase, deoxyribokinase, and
CC nucleoside 2-deoxyribosyltransferase
XX

XX Sequence 239 AA;

Query Match 99.8%; Score 1222; DB 4; Length 239;
Best Local Similarity 99.6%; Pred. No. 2.4e-123;
Matches 238; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MATPHINAEAGDFADVLMPCGDLRAKYIAETFLDAREVNNVRGMLGFTGYKGRKISV 60
DB 1 MATPHINAEAGDFADVLMPCGDLRAKYIAETFLDAREVNNVRGMLGFTGYKGRKISV 60
QY 61 MGHGVGIPSCSYTYTKELITDFGVKKIIRVSGCAVLPHVKLRDVGIGACTDSKVNIR 120
DB 61 MGHGVGIPSCSYTYTKELITDFGVKKIIRVSGCAVLPHVKLRDVGIGACTDSKVNIR 120
QY 121 FKDHDFAAIADFDMVRNAVDAKALGIDARVGNLFSADLFYSPDGMFDMVEKYGILGVE 180
DB 121 FKDHDFAAIADFDMVRNAVDAKALGIDARVGNLFSADLFYSPDGMFDMVEKYGILGVE 180
QY 181 MEAGIYGVAAEFAGAKALTICTVSDHIRTHTQTTAAERQTTFNNDMIKIALESVLLGDKE 239
DB 181 MEAGIYGVAAEFAGAKALTICTVSDHIRTHTQTTAAERQTTFNNDMIKIALESVLLGDKE 239

RESULT 4
AAE35397 standard; protein; 239 AA.

XX AAE35397;
AC AAE35397;
XX
DT 17-JUN-2003 (first entry)
XX
DE Escherichia coli cytosine nucleoside phosphorylase.
XX
KW Cytosine nucleoside phosphorylase; cytosine nucleoside compound;
KW pharmaceutical; enzyme.
XX
OS Escherichia coli.
XX
PN EP1254959-A2.
XX
PD 06-NOV-2002.
XX
PF 01-MAY-2002; 2002EP-00253075.
XX
PR 01-MAY-2001; 2001JP-00134352.
XX
PA (MITA) MITSUI CHEM INC.
XX
XX Araki T, Ikeda I, Matoishi K, Abe R, Oikawa T, Matsuba Y;
PI Ishibaishi H, Nagahara K, Fukui Y;

XX WPI; 2003-158787/16.
DR N-PSDB; AAD54112.
XX

PT Producing cytosine nucleoside compound by reacting sugar phosphate and
PT cytosine or cytosine derivative in the presence of cytosine nucleoside
PT phosphorylase.
XX

PS Claim 11; Page 18; 32pp; English.

XX The invention relates to a method for producing cytosine nucleoside
CC compound by reacting sugar phosphate and cytosine or cytosine derivative
CC in the presence of cytosine nucleoside phosphorylase. The method is used
CC to produce cytosine nucleoside compounds which are intermediates for
CC pharmaceuticals. The present sequence is Escherichia coli cytosine
CC nucleoside phosphorylase
XX

XX Sequence 239 AA;

Query Match 99.8%; Score 1222; DB 6; Length 239;
Best Local Similarity 99.6%; Pred. No. 2.4e-123;
Matches 238; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MATPHINAEAGDFADVLMPCGDLRAKYIAETFLDAREVNNVRGMLGFTGYKGRKISV 60
DB 1 MATPHINAEAGDFADVLMPCGDLRAKYIAETFLDAREVNNVRGMLGFTGYKGRKISV 60
QY 61 MGHGVGIPSCSYTYTKELITDFGVKKIIRVSGCAVLPHVKLRDVGIGACTDSKVNIR 120
DB 61 MGHGVGIPSCSYTYTKELITDFGVKKIIRVSGCAVLPHVKLRDVGIGACTDSKVNIR 120
QY 121 FKDHDFAAIADFDMVRNAVDAKALGIDARVGNLFSADLFYSPDGMFDMVEKYGILGVE 180
DB 121 FKDHDFAAIADFDMVRNAVDAKALGIDARVGNLFSADLFYSPDGMFDMVEKYGILGVE 180
QY 181 MEAGIYGVAAEFAGAKALTICTVSDHIRTHTQTTAAERQTTFNNDMIKIALESVLLGDKE 239
DB 181 MEAGIYGVAAEFAGAKALTICTVSDHIRTHTQTTAAERQTTFNNDMIKIALESVLLGDKE 239

RESULT 5
ADL27842 standard; protein; 239 AA.

XX ADL27842;
AC ADL27842;
XX
DT 20-MAY-2004 (first entry)
XX
DE E coli deod.
XX
KW selection method; yield; pathogen resistance; nutritional quality;
KW stress resistance; plant; transgenic.
XX
OS Escherichia coli.
XX
PN WO2004013333-A2.
XX
PD 12-FEB-2004.
XX
PF 18-JUL-2003; 2003WO-EP007877.
XX
PR 26-JUL-2002; 2002DE-01034287.
XX
PA (BADI) BASF PLANT SCI GMBH.
XX
PI Kock M, Frank M, Badur R;
XX
DR WPI; 2004-157134/15.
DR N-PSDB; ADL27841.
XX
PT Method for transforming plant cells, useful for preparing transgenic
PT plants for e.g. food or production of chemicals, with selection based on
PT suppressing toxic effects of a marker protein.

XX PS Claim 6; Page 140-141; 201pp; German.

XX CC The present invention relates to a method for preparing transformed plant

CC cells or organisms, which comprises first transfecting a population of

CC plant cells, already containing at least one marker protein that can

CC exert a direct or indirect toxic effect, with at least one nucleic acid

CC sequence together with at least one compound able to reduce expression,

CC amount, activity and/or function of the marker protein. Transfected cells

CC that contain the nucleic acid in their genome and, because of the

CC activity of the compound have a growth advantage relative to non-

CC transformed cells are selected, under conditions where the marker protein

CC exerts its toxic effect on non-transformed cells. The method is

CC especially used to produce transgenic plants, especially those having an

CC advantageous phenotype, e.g. better nutritional quality for humans or

CC animals, production of selected chemicals or pharmaceuticals, increased

CC resistance to pathogens or environmental stress or higher yields. The

CC present sequence is a polypeptide shown in the exemplification of the

CC invention.

XX CC

XX SQ Sequence 239 AA;

Query Match 99.8%; Score 1222; DB 8; Length 239;

Best Local Similarity 99.6%; Pred. No. 2.4e-123;

Matches 238; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MATPHINAEMGDFADVLMPCGDLPRAKYIAETFLDAREVNNVRGMLGFTGYGKRKISV 60

Db 1 MATPHINAEMGDFADVLMPCGDLPRAKYIAETFLDAREVNNVRGMLGFTGYGKRKISV 60

Qy 61 MGHGVGIPSCSIYTKELITDVGKKIIRVSGCGAVLPHVKLRDVGVMGACTDSKVNRI 120

Db 61 MGHGVGIPSCSIYTKELITDVGKKIIRVSGCGAVLPHVKLRDVGVMGACTDSKVNRI 120

Qy 121 FKDHDFAAIADFDMVRNAVDAKALGIDARVGNLFSADLFYSPDGMFDMVEKYGILGVE 180

Db 121 FKDHDFAAIADFDMVRNAVDAKALGIDARVGNLFSADLFYSPDGMFDMVEKYGILGVE 180

Qy 181 MEAGIYGVAAEFGAKALTICTVSDHIRTHTHEQTAAERQTTFNDMIKIALESVLLGDKE 239

Db 181 MEAGIYGVAAEFGAKALTICTVSDHIRTHTHEQTAAERQTTFNDMIKIALESVLLGDKE 239

RESULT 6

ABR44456

ID ABR44456 standard; protein; 239 AA.

XX AC ABR44456;

XX CC

XX DT 07-AUG-2003 (first entry)

XX DE Escherichia coli mutant PNP enzyme M65A.

XX CC

XX KW Purine nucleoside phosphorylase; mutein; tumour; PNP; enzyme; cancer;

XX KW leukaemia; glioblastoma; lymphoma; melanoma; carcinoma; gene therapy.

XX CC

XX OS Escherichia coli.

XX OS Synthetic.

XX CC

XX FH Key Location/Qualifiers

XX FT Misc-difference 65

XX FT /note= "wild-type Met replaced with Ala"

XX PN WO2003035012-A2.

XX PD

XX PF 01-MAY-2003.

XX PF 28-OCT-2002; 2002WO-US034626.

XX PR 26-OCT-2001; 2001US-00035300.

XX PA (UABR-) UAB RES FOUND.

XX PA (SOUR) SOUTHERN RES INST.

PA (CORR) CORNELL RES FOUND INC.

XX PI Ealick SE, Parker WB, Seerist JA, Sorscher EJ;

XX DR WPI; 2003-421350/39.

XX PT Novel purified mutant purine cleaving enzyme having different biological

PT activity than a wild-type purine cleaving enzyme, useful for treating a

PT pathological condition characterized by abnormal cell growth.

XX PS Claim 13; Page; 67pp; English.

XX CC The invention relates to mutant purine nucleoside phosphorylase (PNP)

CC enzymes and nucleic acid molecules encoding such enzymes. These mutant

CC enzymes have greater biological activity than wild-type enzymes. PNP

CC enzymes are useful for treating a pathological condition characterised by

CC abnormal cell growth such as that occurring in cancer of the skin,

CC breast, bone, colon, head or neck, kidney, larynx, lung, liver, ovary,

CC nasopharynx, oesophagus, pancreas, prostate, bladder, rectum, stomach,

CC thyroid, testicle and uterus as well as other conditions characterised by

CC abnormal cell growth such as myeloid leukaemia, glioblastoma and

CC lymphoma. Sequences of the invention are useful for treating metastatic

CC solid tumours such as melanoma, pancreatic, liver or colonic carcinoma or

CC to activate prodrugs in the treatment of cancer. They are useful for

CC treating virally infected cells. They are also useful in gene therapy.

CC The present sequence is Escherichia coli mutant PNP enzyme (M65A) amino

CC acid sequence. NOTE: This sequence is given in the specification, but may

CC be created by making the wild-type sequence from the mutant sequence

CC given in SEQ ID 2 (AAE37217) and then creating the desired mutation

XX CC

XX SQ Sequence 239 AA;

Query Match 99.7%; Score 1221; DB 6; Length 239;

Best Local Similarity 99.6%; Pred. No. 3.1e-123;

Matches 238; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MATPHINAEMGDFADVLMPCGDLPRAKYIAETFLDAREVNNVRGMLGFTGYGKRKISV 60

Db 1 MATPHINAEMGDFADVLMPCGDLPRAKYIAETFLDAREVNNVRGMLGFTGYGKRKISV 60

Qy 61 MGHGVGIPSCSIYTKELITDVGKKIIRVSGCGAVLPHVKLRDVGVMGACTDSKVNRI 120

Db 61 MGHGVGIPSCSIYTKELITDVGKKIIRVSGCGAVLPHVKLRDVGVMGACTDSKVNRI 120

Qy 121 FKDHDFAAIADFDMVRNAVDAKALGIDARVGNLFSADLFYSPDGMFDMVEKYGILGVE 180

Db 121 FKDHDFAAIADFDMVRNAVDAKALGIDARVGNLFSADLFYSPDGMFDMVEKYGILGVE 180

Qy 181 MEAGIYGVAAEFGAKALTICTVSDHIRTHTHEQTAAERQTTFNDMIKIALESVLLGDKE 239

Db 181 MEAGIYGVAAEFGAKALTICTVSDHIRTHTHEQTAAERQTTFNDMIKIALESVLLGDKE 239

RESULT 7

AAE35404

ID AAE35404 standard; protein; 239 AA.

XX AC AAE35404;

XX CC

XX DT 17-JUN-2003 (first entry)

XX DE Escherichia coli cytosine nucleoside phosphorylase mutant, V1041.

XX KW Cytosine nucleoside phosphorylase; cytosine nucleoside compound;

XX KW pharmaceutical; enzyme; mutant; mutein.

XX OS Escherichia coli.

XX OS Synthetic.

XX CC

XX FH Key Location/Qualifiers

XX FT Misc-difference 104

XX FT /note= "wild-type Val is substituted with Ile"

XX CC


```
PN EP1254959-A2.
XX
XX PD 06-NOV-2002.
XX
XX PF 01-MAY-2002; 2002EP-00253075.
XX
XX PR 01-MAY-2001; 2001JP-00134352.
XX
XX PA (MITA ) MITSUI CHEM INC.
XX
XX PI Araki T, Ikeda I, Matoishi K, Abe R, Oikawa T, Matsuba Y;
XX Ishibashi H, Nagahara K, Fukui Y;
XX WPI; 2003-158787/16.
XX
XX PT Producing cytosine nucleoside compound by reacting sugar phosphate and
XX cytosine or cytosine derivative in the presence of cytosine nucleoside
XX phosphorylase.
XX
XX PS Example 13; Page; 32pp; English.
XX
XX CC The invention relates to a method for producing cytosine nucleoside
XX compound by reacting sugar phosphate and cytosine or cytosine derivative
XX in the presence of cytosine nucleoside phosphorylase. The method is used
XX to produce cytosine nucleoside compounds which are intermediates for
XX pharmaceuticals. The present sequence is Escherichia coli cytosine
XX nucleoside phosphorylase. Note: This sequence is not shown in the
XX specification, however it is constructed based on E. coli cytosine
XX nucleoside phosphorylase (AAE35397) sequence shown in column 18
XX
XX SQ Sequence 239 AA;
XX
XX Query Match 99.7%; Score 1221; DB 6; Length 239;
XX Best Local Similarity 99.2%; Pred. No. 3.1e-123;
XX Matches 237; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
XX
QY 1 MATPHINAEAGDFADVLMPGDPLRAKYIAETFLDAREVNNVRGMLGFTGYKGRKISV 60
DB 1 MATPHINAEAGDFADVLMPGDPLRAKYIAETFLDAREVNNVRGMLGFTGYKGRKISV 60
QY 61 MGHGVGIPSCSIYTKELITDFGVKKIIRVSGGAVLPHVKLRDVGIGACTDSKVNRI 120
DB 61 MGHGMGLPSCSIYTKELITDFGVKKIIRVSGGAVLPHVKLRDVGIGACTDSKVNRI 120
QY 121 FKDHDFAAIADFDVNRNAVDAAKALGIDARVGNLFSADLFYSPDGEFDMVEKYGILGVE 180
DB 121 FKDHDFAAIADFDVNRNAVDAAKALGIDARVGNLFSADLFYSPDGEFDMVEKYGILGVE 180
QY 181 MEAAGIYGVAEFGAKALTICTVSDHIRTHTQTTAAERQTTFNNDMIKIALESVLLGDKE 239
DB 181 MEAAGIYGVAEFGAKALTICTVSDHIRTHTQTTAAERQTTFNNDMIKIALESVLLGDKE 239
XX
XX RESULT 8
XX AAE35402
XX ID AAE35402 standard; protein; 239 AA.
XX
XX AC AAE35402;
XX
XX DT 17-JUN-2003 (first entry)
XX
XX DE Escherichia coli cytosine nucleoside phosphorylase mutant, I67L.
XX
XX KW Cytosine nucleoside phosphorylase; cytosine nucleoside compound;
XX pharmaceutical; enzyme; mutant; muten.
XX
XX OS Escherichia coli.
XX
XX OS Synthetic.
XX
XX FH Key Location/Qualifiers
XX FT Misc-difference 67
XX FT /note= "wild-type Ile is substituted with Leu"
XX
```

```
PN EP1254959-A2.
XX
XX PD 06-NOV-2002.
XX
XX PF 01-MAY-2002; 2002EP-00253075.
XX
XX PR 01-MAY-2001; 2001JP-00134352.
XX
XX PA (MITA ) MITSUI CHEM INC.
XX
XX PI Araki T, Ikeda I, Matoishi K, Abe R, Oikawa T, Matsuba Y;
XX Ishibashi H, Nagahara K, Fukui Y;
XX WPI; 2003-158787/16.
XX
XX PT Producing cytosine nucleoside compound by reacting sugar phosphate and
XX cytosine or cytosine derivative in the presence of cytosine nucleoside
XX phosphorylase.
XX
XX PS Example 13; Page; 32pp; English.
XX
XX CC The invention relates to a method for producing cytosine nucleoside
XX compound by reacting sugar phosphate and cytosine or cytosine derivative
XX in the presence of cytosine nucleoside phosphorylase. The method is used
XX to produce cytosine nucleoside compounds which are intermediates for
XX pharmaceuticals. The present sequence is Escherichia coli cytosine
XX nucleoside phosphorylase. Note: This sequence is not shown in the
XX specification, however it is constructed based on E. coli cytosine
XX nucleoside phosphorylase (AAE35397) sequence shown in column 18
XX
XX SQ Sequence 239 AA;
XX
XX Query Match 99.6%; Score 1220; DB 6; Length 239;
XX Best Local Similarity 99.2%; Pred. No. 3.9e-123;
XX Matches 237; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
XX
QY 1 MATPHINAEAGDFADVLMPGDPLRAKYIAETFLDAREVNNVRGMLGFTGYKGRKISV 60
DB 1 MATPHINAEAGDFADVLMPGDPLRAKYIAETFLDAREVNNVRGMLGFTGYKGRKISV 60
QY 61 MGHGVGIPSCSIYTKELITDFGVKKIIRVSGGAVLPHVKLRDVGIGACTDSKVNRI 120
DB 61 MGHGMGLPSCSIYTKELITDFGVKKIIRVSGGAVLPHVKLRDVGIGACTDSKVNRI 120
QY 121 FKDHDFAAIADFDVNRNAVDAAKALGIDARVGNLFSADLFYSPDGEFDMVEKYGILGVE 180
DB 121 FKDHDFAAIADFDVNRNAVDAAKALGIDARVGNLFSADLFYSPDGEFDMVEKYGILGVE 180
QY 181 MEAAGIYGVAEFGAKALTICTVSDHIRTHTQTTAAERQTTFNNDMIKIALESVLLGDKE 239
DB 181 MEAAGIYGVAEFGAKALTICTVSDHIRTHTQTTAAERQTTFNNDMIKIALESVLLGDKE 239
XX
XX RESULT 9
XX ABR44462
XX ID ABR44462 standard; protein; 239 AA.
XX
XX AC ABR44462;
XX
XX DT 07-AUG-2003 (first entry)
XX
XX DE Escherichia coli mutant PNP enzyme E180D.
XX
XX KW Purine nucleoside phosphorylase; muten; tumour; PNP; enzyme; cancer;
XX leukaemia; glioblastoma; lymphoma; melanoma; carcinoma; gene therapy.
XX
XX OS Escherichia coli.
XX
XX OS Synthetic.
XX
XX FH Key Location/Qualifiers
XX FT Misc-difference 180
XX FT /note= "wild-type Glu replaced with Asp"
XX
```

PN W02003035012-A2.
PD 01-MAY-2003.
XX 28-OCT-2002; 2002WO-US034626.
XX 26-OCT-2001; 2001US-00035300.
XX (UABR-) UAB RES FOUND.
PA (SOUR) SOUTHERN RES INST.
FA (CORR) CORNELL RES FOUND INC.
XX
PI Ealick SE, Parker WB, Seerist JA, Sorscher EJ;
XX WPI; 2003-421350/39.
DR
XX Novel purified mutant purine cleaving enzyme having different biological
PT activity than a wild-type purine cleaving enzyme, useful for treating a
PT pathological condition characterized by abnormal cell growth.
XX
PS Claim 13; Page; 67pp; English.
XX
CC The invention relates to mutant purine nucleoside phosphorylase (PNP)
CC enzymes and nucleic acid molecules encoding such enzymes. These mutant
CC enzymes have greater biological activity than wild-type enzymes. PNP
CC enzymes are useful for treating a pathological condition characterised by
CC abnormal cell growth such as that occurring in cancer of the skin,
CC breast, bone, colon, head or neck, kidney, larynx, lung, liver, ovary,
CC nasopharynx, oesophagus, pancreas, prostate, bladder, rectum, stomach,
CC thyroid, testicle and uterus as well as other conditions characterised by
CC abnormal cell growth such as myeloid leukaemia, glioblastoma and
CC lymphoma. Sequences of the invention are useful for treating metastatic
CC solid tumours such as melanoma, pancreatic, liver or colonic carcinoma or
CC to activate prodrugs in the treatment of cancer. They are useful for
CC treating virally infected cells. They are also useful in gene therapy.
CC The present sequence is Escherichia coli mutant PNP enzyme (E180D) amino
CC acid sequence. NOTE: This sequence is given in the specification, but may
CC be created by making the wild-type sequence from the mutant sequence
CC given in SEQ ID 2 (AAE37217) and then creating the desired mutation
XX
SQ Sequence 239 AA;
Query Match 99.5%; Score 1219; DB 6; Length 239;
Best Local Similarity 99.2%; Pred. No. 5.1e-123;
Matches 237; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
Qy 1 MATPHINAEMGDFADVLLMPGDPPLRAKYIAETFLDAREVNNVRGMLGFTCTYGRKISV 60
Db 1 MATPHINAEMGDFADVLLMPGDPPLRAKYIAETFLDAREVNNVRGMLGFTCTYGRKISV 60
Qy 61 MGHGVGIPSCSIYTKELITDFGVKKIIRVSGGAVLPHVKLRDVGWACTDSKVNRI 120
Db 61 MGHGVGIPSCSIYTKELITDFGVKKIIRVSGGAVLPHVKLRDVGWACTDSKVNRI 120
Qy 121 FKDHDFAAIADFDMVRNAVDAAKALGIDARVGNLFSADLFYSPDGEFVWNEKYGILGVE 180
Db 121 FKDHDFAAIADFDMVRNAVDAAKALGIDARVGNLFSADLFYSPDGEFVWNEKYGILGVD 180
Qy 181 MEAGIYGVAAEFGAKALITCTVSDHIRTHTQTTAAERQTTFNMDIKIALESVLLGDKE 239
Db 181 MEAGIYGVAAEFGAKALITCTVSDHIRTHTQTTAAERQTTFNMDIKIALESVLLGDKE 239
RESULT 10
ABR44458
ID ABR44458 standard; protein; 239 AA.
XX
AC ABR44458;
XX
DT 07-AUG-2003 (first entry)
XX
DE Escherichia coli mutant PNP enzyme M65Q.
XX

KW Purine nucleoside phosphorylase; mutein; tumour; PNP; enzyme; cancer;
KW leukaemia; glioblastoma; lymphoma; carcinoma; gene therapy.
OS Escherichia coli.
OS Synthetic.
FH Key Location/Qualifiers
XX Misc-difference 65 /note= "wild-type Met replaced with Gln"
FT
XX W02003035012-A2.
XX 01-MAY-2003.
XX 28-OCT-2002; 2002WO-US034626.
XX 26-OCT-2001; 2001US-00035300.
XX (UABR-) UAB RES FOUND.
PA (SOUR) SOUTHERN RES INST.
PA (CORR) CORNELL RES FOUND INC.
XX
XX Ealick SE, Parker WB, Seerist JA, Sorscher EJ;
XX WPI; 2003-421350/39.
DR
XX Novel purified mutant purine cleaving enzyme having different biological
PT activity than a wild-type purine cleaving enzyme, useful for treating a
PT pathological condition characterized by abnormal cell growth.
XX
PS Claim 13; Page; 67pp; English.
XX
CC The invention relates to mutant purine nucleoside phosphorylase (PNP)
CC enzymes and nucleic acid molecules encoding such enzymes. These mutant
CC enzymes have greater biological activity than wild-type enzymes. PNP
CC enzymes are useful for treating a pathological condition characterised by
CC abnormal cell growth such as that occurring in cancer of the skin,
CC breast, bone, colon, head or neck, kidney, larynx, lung, liver, ovary,
CC nasopharynx, oesophagus, pancreas, prostate, bladder, rectum, stomach,
CC thyroid, testicle and uterus as well as other conditions characterised by
CC abnormal cell growth such as myeloid leukaemia, glioblastoma and
CC lymphoma. Sequences of the invention are useful for treating metastatic
CC solid tumours such as melanoma, pancreatic, liver or colonic carcinoma or
CC to activate prodrugs in the treatment of cancer. They are useful for
CC treating virally infected cells. They are also useful in gene therapy.
CC The present sequence is Escherichia coli mutant PNP enzyme (M65Q) amino
CC acid sequence. NOTE: This sequence is given in the specification, but may
CC be created by making the wild-type sequence from the mutant sequence
CC given in SEQ ID 2 (AAE37217) and then creating the desired mutation
XX
SQ Sequence 239 AA;
Query Match 99.5%; Score 1219; DB 6; Length 239;
Best Local Similarity 99.6%; Pred. No. 5.1e-123;
Matches 238; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 1 MATPHINAEMGDFADVLLMPGDPPLRAKYIAETFLDAREVNNVRGMLGFTCTYGRKISV 60
Db 1 MATPHINAEMGDFADVLLMPGDPPLRAKYIAETFLDAREVNNVRGMLGFTCTYGRKISV 60
Qy 61 MGHGVGIPSCSIYTKELITDFGVKKIIRVSGGAVLPHVKLRDVGWACTDSKVNRI 120
Db 61 MGHGVGIPSCSIYTKELITDFGVKKIIRVSGGAVLPHVKLRDVGWACTDSKVNRI 120
Qy 121 FKDHDFAAIADFDMVRNAVDAAKALGIDARVGNLFSADLFYSPDGEFVWNEKYGILGVE 180
Db 121 FKDHDFAAIADFDMVRNAVDAAKALGIDARVGNLFSADLFYSPDGEFVWNEKYGILGVE 180
Qy 181 MEAGIYGVAAEFGAKALITCTVSDHIRTHTQTTAAERQTTFNMDIKIALESVLLGDKE 239
Db 181 MEAGIYGVAAEFGAKALITCTVSDHIRTHTQTTAAERQTTFNMDIKIALESVLLGDKE 239

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 22, 2005, 15:10:32 ; Search time 5506 Seconds
(without alignments)
6336.316 Million cell updates/sec

Title: US-10-035-300A-1

Perfect score: 720
Sequence: 1 atggctaccacacattaa.....tgctggggcgataaagagtaa 720

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.*

1: gb.ba.*

2: gb.htg.*

3: gb.in.*

4: gb.on.*

5: gb.ov.*

6: gb.pat.*

7: gb.ph.*

8: gb.pl.*

9: gb.pr.*

10: gb.ro.*

11: gb.sts.*

12: gb.sy.*

13: gb.un.*

14: gb.vi.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	716.8	99.6	720	6	BD187724 A method
2	716.8	99.6	720	6	CQ774461 Sequence
3	716.8	99.6	720	6	AX087917 Sequence
4	716.8	99.6	720	6	AX590437 Sequence
5	716.8	99.6	1249	1	ECOPNP M60917 E.coli puri
6	716.8	99.6	3031	6	BD261823 Recombina
7	716.8	99.6	3031	6	AX027820 Sequence
8	716.8	99.6	3128	6	BD261824 Recombina
9	716.8	99.6	3128	6	AX027821 Sequence
10	716.8	99.6	3383	6	BD261814 Recombina
11	716.8	99.6	3383	6	AX027811 Sequence
12	716.8	99.6	3934	6	BD261825 Recombina
13	716.8	99.6	3934	6	AX027822 Sequence
14	716.8	99.6	4189	6	BD261816 Recombina
15	716.8	99.6	4189	6	AX027813 Sequence
16	716.8	99.6	5013	6	AR264513 Sequence
17	716.8	99.6	5241	6	BD261818 Recombina
18	716.8	99.6	5241	6	AX027815 Sequence
19	716.8	99.6	5495	6	BD261815 Recombina

20	716.8	99.6	5495	6	AX027812	AX027812 Sequence
21	716.8	99.6	6046	6	BD261826	BD261826 Recombina
22	716.8	99.6	6046	6	AX027823	AX027823 Sequence
23	716.8	99.6	6269	6	BD261820	BD261820 Recombina
24	716.8	99.6	6269	6	AX027817	AX027817 Sequence
25	716.8	99.6	6299	6	BD261821	BD261821 Recombina
26	716.8	99.6	6299	6	AX027818	AX027818 Sequence
27	716.8	99.6	6301	6	BD261817	BD261817 Recombina
28	716.8	99.6	6301	6	AX027814	AX027814 Sequence
29	716.8	99.6	33675	1	U00096_46	Continuation (47 o
30	716.8	99.6	110000	2	EC074221_1	Continuation (2 of
31	716.8	99.6	338534	1	EC0UW93_1	U14003 Escherichia
32	713.6	99.1	720	6	AX087929	AX087929 Sequence
33	705.6	98.0	168764	1	AP002569	AP002569 Escherich
34	705.2	97.9	11599	1	AE005669	AE005669 Escherich
35	699.2	97.1	10535	1	AE015447	AE015447 Shigella
36	699.2	97.1	225944	1	AE016993	AE016993 Shigella
37	688	95.6	86898	1	AE016772	AE016772 Escherich
38	598.4	83.1	21405	1	AE008915	AE008915 Salmonell
39	587.2	81.6	88037	1	AL627284	AL627284 Salmonell
40	587.2	81.6	284233	1	AE016849	AE016849 Salmonell
41	558.4	77.6	732	6	AR384902	AR384902 Sequence
42	527.8	73.3	720	1	KSPNPGNE	X87881 Klebsiella
43	507.2	70.4	110000	1	EX950851_07	Continuation (8 of
44	480	66.7	11093	1	AE013977	AE013977 Yersinia
45	480	66.7	110000	1	EX936398_06	Continuation (7 of

ALIGNMENTS

RESULT 1	BD187724	BD187724	720 bp	DNA	linear	PAT 17-JUL-2003
LOCUS	BD187724	A method of producing a cytosine nucleoside compound.				
DEFINITION	BD187724	A method of producing a cytosine nucleoside compound.				
ACCESSION	BD187724.1	GI:32997463				
VERSION	JP 2003018997-A/3.					
KEYWORDS	Escherichia coli					
SOURCE	Escherichia coli					
ORGANISM	Escherichia coli					
REFERENCE	1 (bases 1 to 720)					
AUTHORS	Araki,T., Ikeda,I., Matoishi,K., Abe,R., Oikawa,T., Matsuba,Y., Nagahara,K., Fukui,Y. and Ishibashi,H.					
TITLE	A method of producing a cytosine nucleoside compound					
JOURNAL	Patent: JP 2003018997-A 3 21-JAN-2003;					
COMMENT	MTSUI CHEMICALS INC					
	OS Escherichia coli					
	PN JP 2003018997-A/3					
	PD 21-JAN-2003					
	PF 01-MAY-2002 JP 2002129867					
	PI TADASHI ARAKI, ICHIRO IKEDA, KAORI MATOISHI, REIKO ABE, TOSHIHIRO OIKAWA,					
	PI YASUKO MATSUBA, KIYOTERU NAGAHARA, YASUSHI FUKUIRI, HIROKI ISHIBASHI					
	PC C12N15/09, C12N1/15, C12N1/19, C12N1/21, C12N5/10, C12N9/10 PC					
	-C12P19/40, C12N15/00,					
	PC C12N5/00					
	CC A method of producing a cytosine nucleoside compound FH					
	Location/Qualifiers					
	FT source					
	1. .720					
	Location/Qualifiers					
	1. .720					
	/organism="Escherichia coli"					
	/mol_type="genomic DNA"					
	/db_xref="taxon:562"					

FEATURES

ORIGIN

Query Match 99.6%; Score 716.8; DB 6; Length 720;
Best Local Similarity 99.7%; Pred. No. 1.6e-181;
Matches 718; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

ORGANISM Escherichia coli
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Escherichia.

REFERENCE 1
AUTHORS Tischer, W., Ihlenfeldt, H.G., Barzu, O., Sakamoto, H., Pistotnik, E.,
Marliere, P. and Pochet, S.
TITLE Enzymatic synthesis of deoxyribonucleosides
JOURNAL Patent: WO 0114566-A 3 01-MAR-2001;
Roche Diagnostics GmbH (DE) ; INSTITUT PASTEUR (FR) ; Pharma-
Waldhof GmbH & Co. KG (DE)

FEATURES
source Location/Qualifiers
1..720
/organism="Escherichia coli"
/mol_type="unassigned DNA"
/db_xref="taxon:562"

CDS
1..720
/note="unnamed protein product"
/codon_start=1
/transl_table=11
/protein_id="CAC34665.1"
/translation="MATPHINAEIMGDFADVVLMPGDPURAKYIAETFLIEDAREVNNVR
GMLGFTGYKGRKISVMGHGIPSCSYTKELITDFGVKKIIRVSGCGAVLPHVKLR
DVVIGMGACTSKVNRIRFKDHPAIIADFMVRNADVDAKALGIDARVGNLFSADLRF
YSPGEMFDVMEKYGILGVEMEAAGIYGVAAEFGAKALTICTVSDHIRTHTQTTAAER
QTTNDMIKIALESVLLGDK"

ORIGIN
Query Match 99.6%; Score 716.8; DB 6; Length 720;
Best Local Similarity 99.7%; Pred. No. 1.6e-181;
Matches 718; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ATGGCTACCCACACATTAATGCAGAAATGGCGGATTTTCGCTGACGTAGTTTGGATGCCA 60
DB 1 ATGGCTACCCACACATTAATGCAGAAATGGCGGATTTTCGCTGACGTAGTTTGGATGCCA 60

QY 61 GCGGACCCGCTCGCTGCGAAGTATATTCGTGAACTTTCTTGAAGATGCCCGTGAAGTG 120
DB 61 GCGGACCCGCTCGCTGCGAAGTATATTCGTGAACTTTCTTGAAGATGCCCGTGAAGTG 120

QY 121 AACACGTTCCGGGTATGCTGGGCTTACCGGTACTTACAAAGCGCGCAAAATTTCCGTA 180
DB 121 AACACGTTCCGGGTATGCTGGGCTTACCGGTACTTACAAAGCGCGCAAAATTTCCGTA 180

QY 181 ATGGGTACCGGTATGGGTATCCCGTCTGCTCCATCTACACAAAGAACTGATCACCGAT 240
DB 181 ATGGGTACCGGTATGGGTATCCCGTCTGCTCCATCTACACAAAGAACTGATCACCGAT 240

QY 241 TTCGGCGTGAAGAAATTAATCCGCGTGGGTTCTGTGGCGCAGTTCTGCGCGACGTAAAA 300
DB 241 TTCGGCGTGAAGAAATTAATCCGCGTGGGTTCTGTGGCGCAGTTCTGCGCGACGTAAAA 300

QY 301 CTGGCGACGCTGTTATCGGTATGGGTGCGTGCACCGATTCCAAGTTAAACGGATCCGT 360
DB 301 CTGGCGACGCTGTTATCGGTATGGGTGCGTGCACCGATTCCAAGTTAAACGGATCCGT 360

QY 361 TTTAAAGACCATGACTTTGCGCGTATCGCTGACTTCGACATGGTGCCTAAGCGAGTAGAT 420
DB 361 TTTAAAGACCATGACTTTGCGCGTATCGCTGACTTCGACATGGTGCCTAAGCGAGTAGAT 420

QY 421 GCAGCTAAAGCACTGGGTATTTGATGCTCGCGTGGGTAACTGTTCTCGCGCTGACCTGTC 480
DB 421 GCAGCTAAAGCACTGGGTATTTGATGCTCGCGTGGGTAACTGTTCTCGCGCTGACCTGTC 480

QY 481 TACTCTCCGGACGGCGAAATGTTCCACGTGATGGGAAATACGGATTCCTGGCGTGAA 540
DB 481 TACTCTCCGGACGGCGAAATGTTCCACGTGATGGGAAATACGGATTCCTGGCGTGAA 540

QY 541 ATGGAAGCGGCTGGTATCTACCGGCTCGCTGCGAGAAATTTGGCGGAAAGCCCTGACCATC 600
DB 541 ATGGAAGCGGCTGGTATCTACCGGCTCGCTGCGAGAAATTTGGCGGAAAGCCCTGACCATC 600

QY 601 TGCACCGTATCTGACCACTCCGCACTACGAGCAGACCACTGCGCGCTGAGCGTCAGACT 660

Db 601 TGCACCGTATCTGACCACTCCGCACTCACAGCAGACCACTGCGCGCTGAGCGTCAGACT 660

QY 661 ACCTTCAACGACATGATCAAAATCGCACTGGAATCCGTTCTGCTGGCGGATAAACAGTAA 720
Db 661 ACCTTCAACGACATGATCAAAATCGCACTGGAATCCGTTCTGCTGGCGGATAAACAGTAA 720

RESULT 4
AX590437
LOCUS AX590437 720 bp DNA linear PAT 27-JAN-2003
DEFINITION Sequence 3 from Patent EP1254959.
ACCESSION AX590437
VERSION AX590437.1 GI:27949070
KEYWORDS
SOURCE Escherichia coli
ORGANISM Escherichia coli
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Escherichia.

REFERENCE 1
AUTHORS Araki, T., Ikeda, I., Matoishi, K., Abe, R., Oikawa, T., Matsuba, Y.,
Ishibashi, H., Nagahara, K. and Fukui, Y.
TITLE Method for producing cytosine nucleoside compounds
JOURNAL Patent: EP 1254959-A 3 06-NOV-2002;
MITSUI CHEMICALS, INC. (JP)

FEATURES
source Location/Qualifiers
1..720
/organism="Escherichia coli"
/mol_type="unassigned DNA"
/db_xref="taxon:562"

ORIGIN
Query Match 99.6%; Score 716.8; DB 6; Length 720;
Best Local Similarity 99.7%; Pred. No. 1.6e-181;
Matches 718; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ATGGCTACCCACACATTAATGCAGAAATGGCGGATTTTCGCTGACGTAGTTTGGATGCCA 60
DB 1 ATGGCTACCCACACATTAATGCAGAAATGGCGGATTTTCGCTGACGTAGTTTGGATGCCA 60

QY 61 GCGGACCCGCTCGCTGCGAAGTATATTCGTGAACTTTCTTGAAGATGCCCGTGAAGTG 120
DB 61 GCGGACCCGCTCGCTGCGAAGTATATTCGTGAACTTTCTTGAAGATGCCCGTGAAGTG 120

QY 121 AACACGTTCCCGGTATGCTGGGCTTACCGGTACTTACAAAGCGCGCAAAATTTCCGTA 180
DB 121 AACACGTTCCCGGTATGCTGGGCTTACCGGTACTTACAAAGCGCGCAAAATTTCCGTA 180

QY 181 ATGGGTACCGGTATGGGTATCCCGTCTGCTCCATCTACACAAAGAACTGATCACCGAT 240
DB 181 ATGGGTACCGGTATGGGTATCCCGTCTGCTCCATCTACACAAAGAACTGATCACCGAT 240

QY 241 TTCGGCGTGAAGAAATTAATCCGCGTGGGTTCTGTGGCGCAGTTCTGCGCGACGTAAAA 300
DB 241 TTCGGCGTGAAGAAATTAATCCGCGTGGGTTCTGTGGCGCAGTTCTGCGCGACGTAAAA 300

QY 301 CTGGCGACGCTGTTATCGGTATGGGTGCGTGCACCGATTCCAAGTTAAACGGATCCGT 360
DB 301 CTGGCGACGCTGTTATCGGTATGGGTGCGTGCACCGATTCCAAGTTAAACGGATCCGT 360

QY 361 TTTAAAGACCATGACTTTGCGCGTATCGCTGACTTCGACATGGTGCCTAAGCGAGTAGAT 420
DB 361 TTTAAAGACCATGACTTTGCGCGTATCGCTGACTTCGACATGGTGCCTAAGCGAGTAGAT 420

QY 421 GCAGCTAAAGCACTGGGTATTTGATGCTCGCGTGGGTAACTGTTCTCGCGCTGACCTGTC 480
DB 421 GCAGCTAAAGCACTGGGTATTTGATGCTCGCGTGGGTAACTGTTCTCGCGCTGACCTGTC 480

QY 481 TACTCTCCGGACGGCGAAATGTTCCACGTGATGGGAAATACGGATTCCTGGCGTGAA 540
DB 481 TACTCTCCGGACGGCGAAATGTTCCACGTGATGGGAAATACGGATTCCTGGCGTGAA 540

QY 541 ATGGAAGCGGCTGGTATCTACCGGCTCGCTGCGAGAAATTTGGCGGAAAGCCCTGACCATC 600
DB 541 ATGGAAGCGGCTGGTATCTACCGGCTCGCTGCGAGAAATTTGGCGGAAAGCCCTGACCATC 600

QY 601 TGCACCGTATCTGACCACTCCGCACTACGAGCAGACCACTGCGCGCTGAGCGTCAGACT 660

```

|||||
541 ATGGAAGCGGTGGTATCTACGGCGTCGTCAGAAATTTGGCGGAAGCCCTGACCATC 600
|||||
601 TGCACCGTATCTGACCATCCGACATCCGACATCCAGACAGACCACTGCGCGTAGCGTCAGACT 660
601 TGCACCGTATCTGACCATCCGACATCCGACATCCAGACAGACCACTGCGCGTAGCGTCAGACT 660
661 ACCTTCAACGACATGATCAAAATCGCACTGGAATCCGTTCTGCTGGCGGATAAAGAGTAA 720
661 ACCTTCAACGACATGATCAAAATCGCACTGGAATCCGTTCTGCTGGCGGATAAAGAGTAA 720

RESULT 5
ECOPNP      1249 bp      DNA      linear      BCT 26-APR-1993
LOCUS      E.coli purine nucleoside phosphorylase (deoD) gene, complete cds.
ACCESSION  M60917
VERSION     M60917.1 GI:147308
KEYWORDS   purine nucleoside phosphorylase.
SOURCE     Escherichia coli
ORGANISM   Escherichia coli
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Escherichia.
Hershfield,M.S., Chaffee,S., Koro-Johnson,L., Mary,A., Smith,A.A.
Use of site-directed mutagenesis to enhance the epitope-shielding
effect of covalent modification of proteins with polyethylene
glycol
JOURNAL    Proc. Natl. Acad. Sci. U.S.A. 88 (16), 7185-7189 (1991)
MEDLINE    91334430
PUBMED     1714590
COMMENT    Original source text: Escherichia coli (strain K-12) DNA.
FEATURES   Location/Qualifiers
            1..1249
            /organism="Escherichia coli"
            /mol_type="genomic DNA"
            /strain="K-12"
            /db_xref="taxon:562"
            111..115
            /note="putative"
            123..842
            /gene="deoD"
            123..842
            /gene="deoD"
            /EC_number="2.4.2.1"
            /note="putative"
            /codon_start=1
            /transl_table=11
            /product="purine nucleoside phosphorylase"
            /protein_id="AA24401.1"
            /db_xref="GI:147309"
            /translation="MATHPHNAEMGDFADVVLMPGDLPRAKYIAETFLDAREVNNVR
            GMLGFTGYGRKISVNHGMGIPSCSIYTKELITDFGVKKIIRVSGCGLVPHVKLR
            DVIWGMCACTDKVNRIRFKDHPAAIDFDMVRNVAADAAKALGIDARVGLFSADLF
            YSPDGMFDMVEKYKILGVMEAGIYVAAEFKAKLTICTVSDHRTHEQTAAER
            OTTENDIKIALESVLGDKE"
            874..901
            /note="putative"
            /function="terminator for deoD transcription"

stem_loop
Query Match      99.6%; Score 716.8; DB 1; Length 1249;
Best Local Similarity 99.7%; Pred. No. 1.7e-181;
Matches 718; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 ATGGCTACCCACACATTAATGCGAATGGCGGATTTTCGCTGACGTAGTTTTCATGCCA 60
Db 123 ATGGCTACCCACACATTAATGCGAATGGCGGATTTTCGCTGACGTAGTTTTCATGCCA 182
Qy 61 GCGGACCCGCTGCGTGCGAAGTATATTCCTGGAACCTTTCTTGAAGATGCCCGTGAAGTG 120
Db 183 GCGGACCCGCTGCGTGCGAAGTATATTCCTGGAACCTTTCTTGAAGATGCCCGTGAAGTG 242

```

```

Qy 121 AACAACTTCCGGGTATGCTGGGCTTTCACCGGTACTTTACAAAGGCGCAAAATTTTCGGTA 180
Db 243 AACAACTTCCGGGTATGCTGGGCTTTCACCGGTACTTTACAAAGGCGCAAAATTTTCGGTA 302
Qy 181 ATGGGTACACGGTGTGGTATCCCGTCTCTGCTCCATCTACACCAAGAACTGATCACCGAT 240
Db 303 ATGGGTACACGGTATGGTATCCCGTCTCTGCTCCATCTACACCAAGAACTGATCACCGAT 362
Qy 241 TTGGGGGTGAAGAAAATTAATCCGCGTGGGTTCTGTGGCGCAGTTTTCGCGCAGCTAAAA 300
Db 363 TTGGGGGTGAAGAAAATTAATCCGCGTGGGTTCTGTGGCGCAGTTTTCGCGCAGCTAAAA 422
Qy 301 CTGGCGACGCTCGTATCGGTATGGGTGCTGCACCGATTTCCAAAGTTAACCCCATCCGT 360
Db 423 CTGGCGACGCTCGTATCGGTATGGGTGCTGCACCGATTTCCAAAGTTAACCCCATCCGT 482
Qy 361 TTTAAAGACCATGACTTTTGCCTATCGCTGACTTCGACATGGTGGCTAAACGAGTAGAT 420
Db 483 TTTAAAGACCATGACTTTTGCCTATCGCTGACTTCGACATGGTGGCTAAACGAGTAGAT 542
Qy 421 GCAGCTTAAGCAGCTGGGTATGATGCTCGGCTGGGTAACTGTTCTCCGCTGACCTGTTTC 480
Db 543 GCAGCTTAAGCAGCTGGGTATGATGCTCGGCTGGGTAACTGTTCTCCGCTGACCTGTTTC 602
Qy 481 TACTCTCCGACGCGCAAAATGTTTCGACGTGATGCAAAAAATACGCGCATTTCTCGGCGTGGAA 540
Db 603 TACTCTCCGACGCGCAAAATGTTTCGACGTGATGCAAAAAATACGCGCATTTCTCGGCGTGGAA 662
Qy 541 ATGGAAGCGGCTGCTATCTACCGCGTGGTGCAGAAATTTTGGCGCGAAAGCCCTGACCATC 600
Db 663 ATGGAAGCGGCTGCTATCTACCGCGTGGTGCAGAAATTTTGGCGCGAAAGCCCTGACCATC 722
Qy 601 TGACCGTATCTGACCATCCGACATCCGACATCCAGACGACACCACTGCGCGTAGCGTCAGACT 660
Db 723 TGACCGGATCTGACCATCCGACATCCGACATCCAGACGACACCACTGCGCGTAGCGTCAGACT 782
Qy 661 ACCTTCAACGACATGATCAAAATCGCACTGGAATCCGTTCTGCTGGCGGATAAAGAGTAA 720
Db 783 ACCTTCAACGACATGATCAAAATCGCACTGGAATCCGTTCTGCTGGCGGATAAAGAGTAA 842

RESULT 6
BD261823    3031 bp      DNA      linear      PAT 17-JUL-2003
LOCUS      Recombinant bacterial strains for the production of natural
DEFINITION nucleosides and modified analogues thereof.
ACCESSION  BD261823
VERSION    BD261823.1 GI:33071591
KEYWORDS   JP 2002533126-A/12.
SOURCE     synthetic construct
ORGANISM   other sequences; artificial sequences.
REFERENCE   1 (bases 1 to 3031)
AUTHORS    Bساتتي,G., Cali,S., Ghisotti,D., Orsini,G., Tonon,G. and Zuffi,G.
TITLES     Recombinant bacterial strains for the production of natural
           nucleosides and modified analogues thereof
JOURNAL    Patent: JP 2002533126-A 12 08-OCT-2002;
           NORPHARMA SPA
COMMENT     OS Artificial Sequence
           PN JP 2002533126-A/12
           PD 08-OCT-2002
           PF 23-DEC-1999 JP 2000591198
           PR 23-DEC-1998 IT MI 98A002792
           PI GIUSEPPINA Bساتتي,SIMONA CALLI,DANIELA GHISOTTI,GAETANO PI
           ORSINI,
           PI GIANCARLO TONON,GABRIELE ZUFFI
           PC C12N15/09,C12N1/21,C12N9/10//C12P19/38,C12P19/40,C12N15/00 CC
           Description of Artificial Sequence: udp and deoD cloned into CC
           pgm746
           CC without upstream ptac promoter
           FH Key Location/Qualifiers
           FT source 1..3031

```

FEATURES	FT	/organism='Artificial Sequence'.									
	source	Location/Qualifiers									
		1. .3031									
		/organism="synthetic construct" /mol_type="genomic DNA" /db_xref="taxon:32630"									
ORIGIN											
	Query Match	99.6%;	Score 716.8;	DB 6;	Length 3031;						
	Best Local Similarity	99.7%;	Pred. No. 1.8e-181;								
	Matches 718;	Conservative 0;	Mismatches 2;	Indels 0;	Gaps 0;						
Qy	1	ATGGCTACCCACACATTAATGCAGAAATGGCGGATTTGCGTAGCTGATGTTTATGATGCCA	60								
Db	101	ATGGCTACCCACACATTAATGCAGAAATGGCGGATTTGCGTAGCTGATGTTTATGATGCCA	160								
Qy	61	GGCGACCCGCTGCGTGGCGAAGTATATTCCTGAACCTTTCCCTTGAAGATGCCCGTGAAGTG	120								
Db	161	GGCGACCCGCTGCGTGGCGAAGTATATTCCTGAACCTTTCCCTTGAAGATGCCCGTGAAGTG	220								
Qy	121	AACAACGTTTCGGGTATGCTCGGCTTCAACCGGTACTTACAAAGGCGCAAAATTTCCGTA	180								
Db	221	AACAACGTTTCGGGTATGCTCGGCTTCAACCGGTACTTACAAAGGCGCAAAATTTCCGTA	280								
Qy	181	ATGGGTACAGGTGTTGGTATCCCGTCTGCTGCTCATCTACACAAAGAACTGATCACCGAT	240								
Db	281	ATGGGTACAGGTATGGGTATCCCGTCTGCTGCTCATCTACACAAAGAACTGATCACCGAT	340								
Qy	241	TTCCGCGTGAAGAAATTTATCCGGTGGGTTCTGTGGCGCAGTTCTGCCGACGTAATAA	300								
Db	341	TTCCGCGTGAAGAAATTTATCCGGTGGGTTCTGTGGCGCAGTTCTGCCGACGTAATAA	400								
Qy	301	CTGGCGCAGCTGCTTATCGGTATGGGTGCTGCACCCGATTCCTCAAGTTAAACCGATCCGT	360								
Db	401	CTGGCGCAGCTGCTTATCGGTATGGGTGCTGCACCCGATTCCTCAAGTTAAACCGATCCGT	460								
Qy	361	TTTAAAGACCATGACTTTGCGGTATCGCTGACCTTGCATGCGTAAACGCGTAGAT	420								
Db	461	TTTAAAGACCATGACTTTGCGGTATCGCTGACCTTGCATGCGTAAACGCGTAGAT	520								
Qy	421	GCAGCTAAGCAGCTGGGTATTTGATGCTCGCGTGGGTAAACCTGTTCTCCGCTGACCTGTC	480								
Db	521	GCAGCTAAGCAGCTGGGTATTTGATGCTCGCGTGGGTAAACCTGTTCTCCGCTGACCTGTC	580								
Qy	481	TACTCTCCGGACGGCGAAATGTTTCGACGTGATGGAATAATACGGCATTTCTCGCGGTGAA	540								
Db	581	TACTCTCCGGACGGCGAAATGTTTCGACGTGATGGAATAATACGGCATTTCTCGCGGTGAA	640								
Qy	541	ATGGAAGCGGTGTTATCTACGGGTGCTGCGAGAAATTTGGCGGAAGCCCTGACCATC	600								
Db	641	ATGGAAGCGGTGTTATCTACGGGTGCTGCGAGAAATTTGGCGGAAGCCCTGACCATC	700								
Qy	601	TGCACCGTATCTGACCATCCGACCTACAGCAGACCACTGCGGCTGAGCGTCAGACT	660								
Db	701	TGCACCGTATCTGACCATCCGACCTACAGCAGACCACTGCGGCTGAGCGTCAGACT	760								
Qy	661	ACCTTCAACGACATGATCAAAATCGCACTGGAATCCGTTCTGCTGGGGGATAAAGAGTAA	720								
Db	761	ACCTTCAACGACATGATCAAAATCGCACTGGAATCCGTTCTGCTGGGGGATAAAGAGTAA	820								
RESULT 7											
AX027820											
LOCUS	3031 bp DNA linear PAT 16-SEP-2000										
DEFINITION	Sequence 12 from Patent WO0039307.										
ACCESSION	AX027820										
VERSION	AX027820.1										
KEYWORDS	GI:10188664										
SOURCE	synthetic construct										
ORGANISM	other sequences; artificial sequences.										
REFERENCE	1										
AUTHORS	Bestetti,G., Cali,S., Orsini,G., Tonon,G., Zuffi,G. and Ghisotti,D.										

TITLE		Recombinant bacterial strains for the production of natural nucleosides and modified analogues thereof	
JOURNAL		Patent: WO 0039307-A 12 06-JUL-2000; BESTETTI GIUSEPPINA (IT) ; CALI SIMONA (IT) ; NORPHARMA SPA (IT) ; ORSINI GAETANO (IT) ; TONON GIANCARLO (IT) ; ZUFFI GABRIELE (IT) ; GHISOTTI DANIELA (IT)	
FEATURES		Location/Qualifiers	
source		1. .3031 /organism="synthetic construct" /mol_type="unassigned DNA" /db_xref="taxon:32630" /note="udp and deod cloned into pGM746 without upstream ptac promoter"	
ORIGIN		Query Match 99.6%; Score 716.8; DB 6; Length 3031; Best Local Similarity 99.7%; Pred. No. 1.8e-181; Matches 718; Conservative 0; Mismatches 2; Indels 0; Gaps 0;	
Qy	1	ATGGCTACCCACACATTAATGCAGAAATGGCGGATTTGCGTAGCTGATGTTTGTATGCCA	60
Db	101	ATGGCTACCCACACATTAATGCAGAAATGGCGGATTTGCGTAGCTGATGTTTGTATGCCA	160
Qy	61	GGCGACCCGCTGCGTGGCGAAGTATATTCGTGAACTTTCTCTTGAAGATGCCCGTGAAGTG	120
Db	161	GGCGACCCGCTGCGTGGCGAAGTATATTCGTGAACTTTCTCTTGAAGATGCCCGTGAAGTG	220
Qy	121	AACAACGTTTCGGGTATGCTGGGCTTCAACCGTACTTACAAAGGCGCAAAATTTCCCGTA	180
Db	221	AACAACGTTTCGGGTATGCTGGGCTTCAACCGTACTTACAAAGGCGCAAAATTTCCCGTA	280
Qy	181	ATGGGTACAGGTGTTGGTATCCCGTCTCTCATCTACACAAAGAACTGATCACCGAT	240
Db	281	ATGGGTACAGGTATGGGTATCCCGTCTCTCATCTACACAAAGAACTGATCACCGAT	340
Qy	241	TTCCGCGTGAAGAAATTTATCCGCGTGGGTTCTGTGGCGCAGTTCTGCCGACGTAATAA	300
Db	341	TTCCGCGTGAAGAAATTTATCCGCGTGGGTTCTGTGGCGCAGTTCTGCCGACGTAATAA	400
Qy	301	CTGGCGCAGCTCGTTATCGGTATGGGTGCTGCACCGATTCCAAAGTTAAACCGCATCCGT	360
Db	401	CTGGCGCAGCTCGTTATCGGTATGGGTGCTGCACCGATTCCAAAGTTAAACCGCATCCGT	460
Qy	361	TTTAAAGACCATGACTTTGCGGTATCGCTGACATTCGACATGGTGGCTAACCGAGTAGAT	420
Db	461	TTTAAAGACCATGACTTTGCGGTATCGCTGACATTCGACATGGTGGCTAACCGAGTAGAT	520
Qy	421	GCAGCTAAGCAGCTGGGTATTTGATGCTCGCGTGGGTAACTGTTCTCCGCTGACCTGTTTC	480
Db	521	GCAGCTAAGCAGCTGGGTATTTGATGCTCGCGTGGGTAACTGTTCTCCGCTGACCTGTTTC	580
Qy	481	TACTCTCCGGACGGCGAAATGTTTCGACGTGATGGAAAAATAACGGCATTCCTCGCGGTGAA	540
Db	581	TACTCTCCGGACGGCGAAATGTTTCGACGTGATGGAAAAATAACGGCATTCCTCGCGGTGAA	640
Qy	541	ATGGAAGCGGCTGGTATCTACGGCGTGGTGCAGAAATTTGGCGCGAAGCCCTGACCATC	600
Db	641	ATGGAAGCGGCTGGTATCTACGGCGTGGTGCAGAAATTTGGCGCGAAGCCCTGACCATC	700
Qy	601	TGCACCGTATCTGACCATCCGACATCAGCAGCAGACCACTGCGCGCTGAGCGTCAGACT	660
Db	701	TGCACCGTATCTGACCATCCGACATCAGCAGCAGACCACTGCGCGCTGAGCGTCAGACT	760
Qy	661	ACCTTCAACGACATGATCAAAATCGCAGCTGGGATCCGTTCTGCTGGCGGATAAAGAGTAA	720
Db	761	ACCTTCAACGACATGATCAAAATCGCAGCTGGGATCCGTTCTGCTGGCGGATAAAGAGTAA	820
RESULT 8			
BD261824			
LOCUS			
DEFINITION			
Recombinant bacterial strains for the production of natural nucleosides and modified analogues thereof.			

BD261824 3128 bp DNA linear PAT 17-JUL-2003
Recombinant bacterial strains for the production of natural
nucleosides and modified analogues thereof.

ACCESSION BD261824
VERSION BD261824.1 GI:33071592
KEYWORDS JP 2002533126-A/13.
SOURCE synthetic construct
ORGANISM synthetic construct
REFERENCE 1 (bases 1 to 3128)
AUTHORS Bestetti,G., Cali,S., Ghisotti,D., Orsini,G., Tonon,G. and Zuffi,G.
TITLE Recombinant bacterial strains for the production of natural nucleosides and modified analogues thereof
JOURNAL Patent: JP 2002533126-A 13 08-OCT-2002;
NORPHARMA SPA
COMMENT OS Artificial Sequence
PN JP 2002533126-A/13
PD 08-OCT-2002
PF 23-DEC-1999 JP 2000591198
PR 23-DEC-1998 IT MI 98A002792
PI GIUSEPPINA BESTETTI,SIMONA CALI,DANIELA GHISOTTI,GAETANO PI ORSINI,
PI GIANCARLO TONON,GABRIELE ZUFFI
PC C12N15/09,C12N1/21,C12N9/10//C12P19/38,C12P19/40,C12N15/00 CC
Description of Artificial Sequence: deoD cloned downstream CC
ptac promoter
FH Key Location/Qualifiers
FT source 1..3128
FT /organism='Artificial Sequence'.
FEATURES
source Location/Qualifiers
1..3128
/organism="synthetic construct"
/mol_type="genomic DNA"
/db_xref="taxon:32630"
ORIGIN
Query Match 99.6%; Score 716.8; DB 6; Length 3128;
Best Local Similarity 99.7%; Pred. No. 1.8e-181;
Matches 718; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 ATGGCTACCCACACATTAATGCAGAAATGGCGATTTCGCTGACGTAGTATTGATGCCA 60
DB 198 ATGGCTACCCACACATTAATGCAGAAATGGCGATTTCGCTGACGTAGTATTGATGCCA 257
QY 61 GGCACCCGCTGCTGCGAAGTATATTGCTGAAACTTTCTTGAAGATGCCGGAAGTG 120
DB 258 GGCACCCGCTGCTGCGAAGTATATTGCTGAAACTTTCTTGAAGATGCCGGAAGTG 317
QY 121 AACAACTTCCGCGTATGCTGGGCTTACCGGTACTTACAAAGCCCGCAAAATTCGCGTA 180
DB 318 AACAACTTCCGCGTATGCTGGGCTTACCGGTACTTACAAAGCCCGCAAAATTCGCGTA 377
QY 181 ATGGGTACCGGTATGCTATCCCGTCTCTGCTCCATCTACACCAAGAACTGATCACCGAT 240
DB 378 ATGGGTACCGGTATGCGGTATCCCGTCTCTGCTCCATCTACACCAAGAACTGATCACCGAT 437
QY 241 TTCCGCGTGAAGAAAATTAATCCGCTGGGTTCTGTGCGCAGTTCTGCGCAGGTAAAA 300
DB 438 TTCCGCGTGAAGAAAATTAATCCGCTGGGTTCTGTGCGCAGTTCTGCGCAGGTAAAA 497
QY 301 CTGCGCACGTCGTATCGGTATGGGTGCTGACCGATTCCAAAGTTAACCGCATCCGT 360
DB 498 CTGCGCACGTCGTATCGGTATGGGTGCTGACCGATTCCAAAGTTAACCGCATCCGT 557
QY 361 TTTAAAGACCATGACTTTGCGGCTATCGCTGACTTCGACATGGTGGCGAGTAGAT 420
DB 558 TTTAAAGACCATGACTTTGCGGCTATCGCTGACTTCGACATGGTGGCGAGTAGAT 617
QY 421 GCAGCTAAAGCAGTGGGTATTGATGCTTCGGGTGGGTAACTGTTCTCCGCTGACCTTTC 480
DB 618 GCAGCTAAAGCAGTGGGTATTGATGCTTCGGGTGGGTAACTGTTCTCCGCTGACCTTTC 677
QY 481 TACTCTCGGACGCGGAATGTTGAGCTGATGAAAAATACGGCATTCCTCGGCTGGAA 540
DB 678 TACTCTCGGACGCGGAATGTTGAGCTGATGAAAAATACGGCATTCCTCGGCTGGAA 737

QY 541 ATGGAAGCGCTCGTATCTACGCGCTGCTGCAGAAATTTTGGCGGAAGCCCTGACCATC 600
DB 738 ATGGAAGCGCTCGTATCTACGCGCTGCTGCAGAAATTTTGGCGGAAGCCCTGACCATC 797
QY 601 TGCACCGTATCTGACACCATCCGCACTCAGGACGACCACTGCCGTGAGCGTCAGACT 660
DB 798 TGCACCGTATCTGACCACTCCGCACTCAGGACGACCACTGCCGTGAGCGTCAGACT 857
QY 661 ACCTTCAACGACATGATCAAAATCGCACTGGAATCCGTTCTGCTGGCGGATAAAGAGTAA 720
DB 858 ACCTTCAACGACATGATCAAAATCGCACTGGAATCCGTTCTGCTGGCGGATAAAGAGTAA 917
RESULT 9
AX027821
LOCUS AX027821 3128 bp DNA linear PAT 16-SEP-2000
DEFINITION Sequence 13 from Patent WO039307.
ACCESSION AX027821
VERSION AX027821.1 GI:10188665
KEYWORDS synthetic construct
SOURCE synthetic construct
ORGANISM other sequences; artificial sequences.
REFERENCE 1
AUTHORS Bestetti,G., Cali,S., Orsini,G., Tonon,G., Zuffi,G. and Ghisotti,D.
TITLE Recombinant bacterial strains for the production of natural nucleosides and modified analogues thereof
JOURNAL Patent: WO 0039307-A 13 06-JUL-2000;
BESTETTI GIUSEPPINA (IT) ; CALI SIMONA (IT) ; NORPHARMA SPA (IT) ; ORSINI GAETANO (IT) ; TONON GIANCARLO (IT) ; ZUFFI GABRIELE (IT) ; GHISOTTI DANIELA (IT)
FEATURES
source Location/Qualifiers
1..3128
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/notes="deoD cloned downstream ptac promoter"
ORIGIN
Query Match 99.6%; Score 716.8; DB 6; Length 3128;
Best Local Similarity 99.7%; Pred. No. 1.8e-181;
Matches 718; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 ATGGCTACCCACACATTAATGCAGAAATGGCGATTTCGCTGACGTAGTATTGATGCCA 60
DB 198 ATGGCTACCCACACATTAATGCAGAAATGGCGATTTCGCTGACGTAGTATTGATGCCA 257
QY 61 GGCACCCGCTGCTGCGAAGTATATTGCTGAAACTTTCTTGAAGATGCCGGAAGTG 120
DB 258 GGCACCCGCTGCTGCGAAGTATATTGCTGAAACTTTCTTGAAGATGCCGGAAGTG 317
QY 121 AACAACTTCCGCGTATGCTGGGCTTACCGGTACTTACAAAGCCCGCAAAATTCGCGTA 180
DB 318 AACAACTTCCGCGTATGCTGGGCTTACCGGTACTTACAAAGCCCGCAAAATTCGCGTA 377
QY 181 ATGGGTACCGGTATGCTATCCCGTCTCTGCTCCATCTACACCAAGAACTGATCACCGAT 240
DB 378 ATGGGTACCGGTATGCGGTATCCCGTCTCTGCTCCATCTACACCAAGAACTGATCACCGAT 437
QY 241 TTCCGCGTGAAGAAAATTAATCCGCTGGGTTCTGTGCGCAGTTCTGCGCAGGTAAAA 300
DB 438 TTCCGCGTGAAGAAAATTAATCCGCTGGGTTCTGTGCGCAGTTCTGCGCAGGTAAAA 497
QY 301 CTGCGCACGTCGTATCGGTATGGGTGCTGACCGATTCCAAAGTTAACCGCATCCGT 360
DB 498 CTGCGCACGTCGTATCGGTATGGGTGCTGACCGATTCCAAAGTTAACCGCATCCGT 557
QY 361 TTTAAAGACCATGACTTTGCGGCTATCGCTGACTTCGACATGGTGGCGAGTAGAT 420
DB 558 TTTAAAGACCATGACTTTGCGGCTATCGCTGACTTCGACATGGTGGCGAGTAGAT 617
QY 421 GCAGCTAAAGCAGTGGGTATTGATGCTTCGGGTGGGTAACTGTTCTCCGCTGACCTTTC 480
DB 678 TACTCTCGGACGCGGAATGTTGAGCTGATGAAAAATACGGCATTCCTCGGCTGGAA 737

```
Db      618 GCAGCTAAGCAGCTGGGTATTGATGCTCGCGTGGGTAACTGTTCTCGCTGACCTGTTTC 677
Qy      481 TACTCTCGGACGGCGAAATGTTTCGACGTGATGGAATAATACGGCAATTCCTCGGCGTGGAA 540
Db      678 TACTCTCGGACGGCGAAATGTTTCGACGTGATGGAATAATACGGCAATTCCTCGGCGTGGAA 737
Qy      541 ATGAAGCGGCTGGTATCTACGGCGTCTCGCTGAGAAATTTGGCGGGAAGCCCTGACCATC 600
Db      738 ATGAAGCGGCTGGTATCTACGGCGTCTCGCTGAGAAATTTGGCGGGAAGCCCTGACCATC 797
Qy      601 TGCACCGTATCTGACCATCCGCACTCACCAGCAGACCACTGCGCGTGGCGTGCAGACT 660
Db      798 TGCACCGTATCTGACCATCCGCACTCACCAGCAGACCACTGCGCGTGGCGTGCAGACT 857
Qy      661 ACCTTCAACGACATGATCAAAATCGCACTGGAATCCGTTCTGCTGGCGGATAAAGAGTAA 720
Db      858 ACCTTCAACGACATGATCAAAATCGCACTGGAATCCGTTCTGCTGGCGGATAAAGAGTAA 917

RESULT 10
BD261814
LOCUS   BD261814          3383 bp    DNA        linear    PAT 17-JUL-2003
DEFINITION Recombinant bacterial strains for the production of natural
nucleosides and modified analogues thereof.
ACCESSION BD261814
VERSION   BD261814.1 GI:33071582
KEYWORDS JP 2002533126-A/3.
SOURCE   synthetic construct
ORGANISM other sequences; artificial sequences.
REFERENCE 1 (bases 1 to 3383)
AUTHORS Bestetti,G., Cali,S., Ghisotti,D., Orsini,G., Tonon,G. and Zuffi,G.
TITLE    Recombinant bacterial strains for the production of natural
nucleosides and modified analogues thereof
JOURNAL  Patent: JP 2002533126-A 3 08-OCT-2002;
NORPHARMA SPA
COMMENT  OS Artificial Sequence
PN JP 2002533126-A/3
PD 08-OCT-2002
PR 23-DEC-1999 JP 2000591198
PR 23-DEC-1998 IT MI 98A002792
PI GIUSEPPINA BESTETTI, SIMONA CALI, DANIELA GHISOTTI, GAETANO PI
ORSINI,
PI GIANCARLO TONON, GABRIELE ZUFFI
PC C12N15/09, C12N1/21, C12N9/10//C12P19/38, C12P19/40, C12N15/00 CC
Description of Artificial Sequence: Plasmid
CC deOD
FH Key Location/Qualifiers
FT gene (231)..(960).
FEATURES
source 1. .3383
location/qualifiers
/organism="synthetic construct"
/mol_type="genomic DNA"
/db_xref="taxon:32630"
ORIGIN
Query Match 99.6%; Score 716.8; DB 6; Length 3383;
Best Local Similarity 99.7%; Pred. No. 1.8e-181;
Matches 718; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Qy 1 ATGCTACCCACACATTAATGCGAAATGGCGGATTCGCTGACGTAGTTTGTATGCCA 60
Db 240 ATGGCTACCCACACATTAATGCGAAATGGCGGATTCGCTGACGTAGTTTGTATGCCA 299
Qy 61 GCGACCGCTGCGTGGCAAGTATATGCTGAACTTCCCTGAAGATGCCCGTGAAGTG 120
Db 300 GCGACCGCTGCGTGGCAAGTATATGCTGAACTTCCCTGAAGATGCCCGTGAAGTG 359
Qy 121 AACCAACGTTCCCGGTATGCTGGGCTTACCGGTACTTACAAAGCGCGCAAAATTTCCGTA 180
Db 360 AACCAACGTTCCCGGTATGCTGGGCTTACCGGTACTTACAAAGCGCGCAAAATTTCCGTA 419
Qy 181 ATGGGTACCGGTGTTGGTATCCCGTCTGCTCCATCTACACCAAGAACTGATCACCGAT 240
```

```
Db      420 ATGGGTACCGGTATGGGTATCCCGTCTGCTCCATCTACACCAAGAACTGATCACCGAT 479
Qy      241 TTCCGCGTGAAGAAATTTATCCGCGTGGTTCCTGTGTGGCGCAGTTCTGCCGCACGTAATA 300
Db      480 TTCCGCGTGAAGAAATTTATCCGCGTGGTTCCTGTGTGGCGCAGTTCTGCCGCACGTAATA 539
Qy      301 CTGCGCACTGCTGTTATCGGTATCGGTGCTGCAACCGATTCCTCAAGTTAACCGCATCGT 360
Db      540 CTGCGCACTGCTGTTATCGGTATCGGTGCTGCAACCGATTCCTCAAGTTAACCGCATCGT 599
Qy      361 TTTHAAGACCATGATTTTGGCGCTATCGCTGACTTCGACATGGTGGCTTAACGCACTAGAT 420
Db      600 TTTHAAGACCATGATTTTGGCGCTATCGCTGACTTCGACATGGTGGCTTAACGCACTAGAT 659
Qy      421 GCAGCTAAAGCAGCTGGGTATTGATGCTCGCGTGGGTAACTGTTCTCGCTGACTGTTTC 480
Db      660 GCAGCTAAAGCAGCTGGGTATTGATGCTCGCGTGGGTAACTGTTCTCGCTGACTGTTTC 719
Qy      481 TACTCTCCGACGGCGAAATGTTTCGACGTGATGGAATAATACGGCATTCCTCGGCGTGGAA 540
Db      720 TACTCTCCGACGGCGAAATGTTTCGACGTGATGGAATAATACGGCATTCCTCGGCGTGGAA 779
Qy      541 ATGCAAGCGCTGCTATCTACGGCGTGGTGCAGAAATTTGGCGGAAAGCCCTGACCATC 600
Db      780 ATGCAAGCGCTGCTATCTACGGCGTGGTGCAGAAATTTGGCGGAAAGCCCTGACCATC 839
Qy      601 TGCACCGTATCTGACCATCCGCACTCACCAGCAGACCACTGCGCGTGGCGTGCAGACT 660
Db      840 TGCACCGTATCTGACCATCCGCACTCACCAGCAGACCACTGCGCGTGGCGTGCAGACT 899
Qy      661 ACCTTCAACGACATGATCAAAATCGCACTGGAATCCGTTCTGCTGGCGGATAAAGAGTAA 720
Db      900 ACCTTCAACGACATGATCAAAATCGCACTGGAATCCGTTCTGCTGGCGGATAAAGAGTAA 959

RESULT 11
AX027811
LOCUS   AX027811          3383 bp    DNA        linear    PAT 16-SEP-2000
DEFINITION Sequence 3 from Patent WO0039307.
ACCESSION AX027811
VERSION   AX027811.1 GI:10188655
KEYWORDS synthetic construct
SOURCE    other sequences; artificial sequences.
ORGANISM 1
REFERENCE Bestetti,G., Cali,S., Orsini,G., Tonon,G., Zuffi,G. and Ghisotti,D.
AUTHORS Recombinant bacterial strains for the production of natural
TITLE    nucleosides and modified analogues thereof
JOURNAL  Patent: WO 0039307-A 3 06-JUL-2000;
BESTETTI GIUSEPPINA (IT) ; CALI SIMONA (IT) ; NORPHARMA SPA (IT) ;
ORSINI GAETANO (IT) ; TONON GIANCARLO (IT) ; ZUFFI GABRIELE (IT) ;
GHISOTTI DANIELA (IT)
FEATURES
source 1. .3383
location/qualifiers
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/notice="Plasmid"
ORIGIN
Query Match 99.6%; Score 716.8; DB 6; Length 3383;
Best Local Similarity 99.7%; Pred. No. 1.8e-181;
Matches 718; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Qy 1 ATGCTACCCACACATTAATGCGAAATGGCGGATTCGCTGACGTAGTTTGTATGCCA 60
Db 240 ATGGCTACCCACACATTAATGCGAAATGGCGGATTCGCTGACGTAGTTTGTATGCCA 299
Qy 61 GCGACCGCTGCGTGGCAAGTATATGCTGAACTTCCCTGAAGATGCCCGTGAAGTG 120
Db 300 GCGACCGCTGCGTGGCAAGTATATGCTGAACTTCCCTGAAGATGCCCGTGAAGTG 359
```



```
QY 121 AACAACTTCGGGTATGCTGGGCTTCAACGGTACTTACAAAGCGCGCAAAATTTCCGTA 180
Db 360 AACAACTTCGGGTATGCTGGGCTTCAACGGTACTTACAAAGCGCGCAAAATTTCCGTA 419
QY 181 ATGGGTCAACGGTATGGGTATCCCGTCTGCTGCTTCACTACACCAAGAACTGATCACCGAT 240
Db 420 ATGGGTCAACGGTATGGGTATCCCGTCTGCTGCTTCACTACACCAAGAACTGATCACCGAT 479
QY 241 TTCCGGCTGAAGAAATATTCGGGTATCCCGTCTGCTGCTTCACTACACCAAGAACTGATCACCGAT 300
Db 480 TTCCGGCTGAAGAAATATTCGGGTATCCCGTCTGCTGCTTCACTACACCAAGAACTGATCACCGAT 539
QY 301 CTGGCGACGCTGCTATCGGTATGGGTATCCCGTCTGCTGCTTCACTACACCAAGAACTGATCACCGAT 360
Db 540 CTGGCGACGCTGCTATCGGTATGGGTATCCCGTCTGCTGCTTCACTACACCAAGAACTGATCACCGAT 599
QY 361 TTTAAAGACCATGACTTTGGCGCTATCGCTGACTTCGACATGCTGCTAAGCGAGTAGAT 420
Db 600 TTTAAAGACCATGACTTTGGCGCTATCGCTGACTTCGACATGCTGCTAAGCGAGTAGAT 659
QY 421 GCAGCTAAAGCACTGGGTATTTGATGCTCGGTGGGTAACTGTTCTCCGCTGACCTGTTT 480
Db 660 GCAGCTAAAGCACTGGGTATTTGATGCTCGGTGGGTAACTGTTCTCCGCTGACCTGTTT 719
QY 481 TACTCTCCGACGCGGAAATGCTTCGAGCTGATGGAATAACCGCATTTCTCGCGTGGAA 540
Db 720 TACTCTCCGACGCGGAAATGCTTCGAGCTGATGGAATAACCGCATTTCTCGCGTGGAA 779
QY 541 ATGGAAGCGGTGCTGATCTACGGGCTCGCTGCGAATTTGGCGGAAAGCCCTGACCATC 600
Db 780 ATGGAAGCGGTGCTGATCTACGGGCTCGCTGCGAATTTGGCGGAAAGCCCTGACCATC 839
QY 601 TGCACCGTATCTGACCACTCCGCACTCACAGAGAGACCACTGCGCGCTGAGCGTACAGT 660
Db 840 TGCACCGTATCTGACCACTCCGCACTCACAGAGAGACCACTGCGCGCTGAGCGTACAGT 899
QY 661 ACCTTCAACGACATGATCAAAATCGCACTGGAATCCGTTCTGCTGGCGGATAAAGAGTAA 720
Db 900 ACCTTCAACGACATGATCAAAATCGCACTGGAATCCGTTCTGCTGGCGGATAAAGAGTAA 959

RESULT 12
BD261825 3934 bp DNA linear PAT 17-JUL-2003
LOCUS Recombinant bacterial strains for the production of natural
DEFINITION nucleosides and modified analogues thereof.
ACCESSION BD261825
VERSION 1 (bases 1 to 3934)
KEYWORDS JP 2002533126-A/14.
SOURCE synthetic construct
ORGANISM
other sequences; artificial sequences.
REFERENCE
1 (bases 1 to 3934)
AUTHORS Bestetti,G., Cali,S., Ghisotti,D., Orsini,G., Tonon,G. and Zuffi,G.
TITLE Recombinant bacterial strains for the production of natural
nucleosides and modified analogues thereof
JOURNAL Patent: JP 2002533126-A 14 08-OCT-2002;
NORPHARMA SPA
COMMENT
OS Artificial Sequence
PN JP 2002533126-A/14
PD 08-OCT-2002
PF 23-DEC-1999 JP 2000591198
PR 23-DEC-1998 IT MI 98A002792
PI GIUSEPPINA BESTETTI,SIMONA CALI,DANIELA GHISOTTI,GAETANO PI
ORSINI,
PI GIANCARLO TONON,GABRIELE ZUFFI
PC C12N15/09,C12N1/21,C12N9/10//C12P19/38,C12P19/40,C12N15/00 CC
Description of Artificial Sequence: udp and deod cloned CC
downstream plac
CC promoter
FH Key Location/Qualifiers
FT source 1..3934
```

```
FEATURES             Location/Qualifiers
     source            1..3934
                     /organism="Artificial Sequence".
ORIGIN
Query Match          99.6%; Score 716.8; DB 6; Length 3934;
Best Local Similarity 99.7%; Pred. No. 1.8e-181;
Matches 718; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 ATGGCTACCCACACACATTAAATGAGAAATGGGCGATTTTCGCTGACGCTAGTCTTTGATGCCA 60
Db 198 ATGGCTACCCACACACATTAAATGAGAAATGGGCGATTTTCGCTGACGCTAGTCTTTGATGCCA 257
QY 61 GGCAGCCCGCTGCTGGGAAGTATATTGCTGAAACTTTCTTGAAGATGCCCGTGAAGTG 120
Db 258 GGCAGCCCGCTGCTGGGAAGTATATTGCTGAAACTTTCTTGAAGATGCCCGTGAAGTG 317
QY 121 AACAGGTTTCGGCGTATGCTGGGCTTCACGGTACTTACAAAGCGCGCAAAATTTCCGTA 180
Db 318 AACAGGTTTCGGCGTATGCTGGGCTTCACGGTACTTACAAAGCGCGCAAAATTTCCGTA 377
QY 181 ATGGGTACACGCTGCTGGTATCCCGTCTGCTGCTTACACCAAGAACTGATCACCGAT 240
Db 378 ATGGGTACACGCTGCTGGTATCCCGTCTGCTTACACCAAGAACTGATCACCGAT 437
QY 241 TTCCGGCTGAAGAAATATTTATCCCGGTGGGTTCCTGTGGCGCAGTTCTGCGCGCAGTAAAA 300
Db 438 TTCCGGCTGAAGAAATATTTATCCCGGTGGGTTCCTGTGGCGCAGTTCTGCGCGCAGTAAAA 497
QY 301 CTGGCGACGCTGCTTATCGGTATGGGTATGGGTGCTGACCGATTTCCAAAGTTAACCGCATCCGT 360
Db 498 CTGGCGACGCTGCTTATCGGTATGGGTGCTGACCGATTTCCAAAGTTAACCGCATCCGT 557
QY 361 TTTAAAGACCATGACTTTGCGGCTATCGCTGACTTCGACATGCTGCGTAAACGAGTAGAT 420
Db 558 TTTAAAGACCATGACTTTGCGGCTATCGCTGACTTCGACATGCTGCGTAAACGAGTAGAT 617
QY 421 GCAGCTAAAGCACTGGGTATTTGATGCTGCGGTGGGTAACTGTTCTCCGCTGACCTGTTT 480
Db 618 GCAGCTAAAGCACTGGGTATTTGATGCTGCGGTGGGTAACTGTTCTCCGCTGACCTGTTT 677
QY 481 TACTCTCCGACGCGGAAATGTTTCGAGCTGATGGAATAACCGCATTTCTCGCGTGGAA 540
Db 678 TACTCTCCGACGCGGAAATGTTTCGAGCTGATGGAATAACCGCATTTCTCGCGTGGAA 737
QY 541 ATGGAAGCGGCTGCTGATCTACGGGCTGCTGCTGCGAATTTGGCGGAAAGCCCTGACCATC 600
Db 738 ATGGAAGCGGCTGCTGATCTACGGGCTGCTGCGAATTTGGCGGAAAGCCCTGACCATC 797
QY 601 TGCACCGTATCTGACCACTCCGCACTCACAGAGAGACCACTGCGCGCTGAGCGTACAGT 660
Db 798 TGCACCGTATCTGACCACTCCGCACTCACAGAGAGACCACTGCGCGCTGAGCGTACAGT 857
QY 661 ACCTTCAACGACATGATCAAAATCGCACTGGAATCCGTTCTGCTGGCGGATAAAGAGTAA 720
Db 858 ACCTTCAACGACATGATCAAAATCGCACTGGAATCCGTTCTGCTGGCGGATAAAGAGTAA 917

RESULT 13
AX027822 3934 bp DNA linear PAT 16-SEP-2000
LOCUS AX027822
DEFINITION Sequence 14 from Patent WO0039307.
ACCESSION AX027822
VERSION AX027822.1 GI:10188666
KEYWORDS
SOURCE synthetic construct
ORGANISM other sequences; artificial sequences.
REFERENCE
1
AUTHORS Bestetti,G., Cali,S., Orsini,G., Tonon,G., Zuffi,G. and Ghisotti,D.
```


TITLE Recombinant bacterial strains for the production of natural nucleosides and modified analogues thereof
JOURNAL Patent: WO 0039307-A 14 06-JUL-2000;
BESTETTI GIUSEPPINA (IT); CALI SIMONA (IT); NORPHARMA SPA (IT); ORSINI GAETANO (IT); TONON GIANCARLO (IT); ZUFFI GABRIELE (IT); GHISOTTI DANIELA (IT)
FEATURES Location/Qualifiers
source 1..3934
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/note="udp and deod cloned downstream ptac promoter"
ORIGIN
Query Match 99.6%; Score 716.8; DB 6; Length 3934;
Best Local Similarity 99.7%; Pred. No. 1.8e-181;
Matches 718; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Qy 1 ATGGCTACCCACACATTAATGCAGAAATGGCGGATTCGCTGACGTAGTTTGGATGCCA 60
Db 198 ATGGCTACCCACACATTAATGCAGAAATGGCGGATTCGCTGACGTAGTTTGGATGCCA 257
Qy 61 GGGGACCCGCTGCGTGGGAGGTATATGCTCAAACTTCTTGAAGATGCCCGTGAAGTG 120
Db 258 GGGGACCCGCTGCGTGGGAGGTATATGCTCAAACTTCTTGAAGATGCCCGTGAAGTG 317
Qy 121 AACAACTGTCGCGGTATGCTGGGCTTCAACCGGTACTTACAAAGGCGCAAAATTTCCGTA 180
Db 318 AACAACTGTCGCGGTATGCTGGGCTTCAACCGGTACTTACAAAGGCGCAAAATTTCCGTA 377
Qy 181 ATGGGTACCGGTGTGGGTATCCCGTCTGCTGCATCTACACAAAGAACTGATCACCAGT 240
Db 378 ATGGGTACCGGTATGGGTATCCCGTCTGCTGCATCTACACAAAGAACTGATCACCAGT 437
Qy 241 TTCCGCGTGAAGAAATATCCGCGTGGTTCCTGTCGCGAGTTTCCCGCAGTAAAA 300
Db 438 TTCCGCGTGAAGAAATATCCGCGTGGTTCCTGTCGCGAGTTTCCCGCAGTAAAA 497
Qy 301 CTGCGCGACGTCGTTATCGGTATGGGTGCTGCACCGATTCCTCAAAAGTTAAACCGATCCGT 360
Db 498 CTGCGCGACGTCGTTATCGGTATGGGTGCTGCACCGATTCCTCAAAAGTTAAACCGATCCGT 557
Qy 361 TTTAAAGACCATGACTTTGCGCGTATCGCTGACTTCGACATGCTGCGTGAAGAGTAGAT 420
Db 558 TTTAAAGACCATGACTTTGCGCGTATCGCTGACTTCGACATGCTGCGTGAAGAGTAGAT 617
Qy 421 GCAGCTAAAGCAGTGGGTATGATGCTGCGTGGGTAACTGTTCTCGCTGACCTGTTTC 480
Db 618 GCAGCTAAAGCAGTGGGTATGATGCTGCGTGGGTAACTGTTCTCGCTGACCTGTTTC 677
Qy 481 TACTCTCCGGACGGCGAAATGTTTCGACGTGATGGAAATAACGGCATTTCTCGGCGTGGAA 540
Db 678 TACTCTCCGGACGGCGAAATGTTTCGACGTGATGGAAATAACGGCATTTCTCGGCGTGGAA 737
Qy 541 ATGAAGCGGTGTTATCTACGGCGTGGTCTGAGAAATTTGGCGGAAAGCCCTGACCATC 600
Db 738 ATGAAGCGGTGTTATCTACGGCGTGGTCTGAGAAATTTGGCGGAAAGCCCTGACCATC 797
Qy 601 TGCACCGTATCTGACCATCGCACTCAGCAGCAGACCACTGCGCTGACGTCAGACT 660
Db 798 TGCACCGTATCTGACCATCGCACTCAGCAGCAGACCACTGCGCTGACGTCAGACT 857
Qy 661 ACCTTTCAACGACATGATCAAAATCGCACTGGAATCCGTTCTGCTGGGGGATAAAGAGTAA 720
Db 858 ACCTTTCAACGACATGATCAAAATCGCACTGGAATCCGTTCTGCTGGGGGATAAAGAGTAA 917
RESULT 14
BD261816 4189 bp DNA linear PAT 17-JUL-2003
LOCUS
DEFINITION Recombinant bacterial strains for the production of natural nucleosides and modified analogues thereof.
ACCESSION BD261816

BD261816.1 GI:33071584
JP 2002533126-A/5.
SOURCE synthetic construct
ORGANISM synthetic construct
REFERENCE 1 (bases 1 to 4189)
Betetti,G., Cali,S., Ghisotti,D., Orsini,G., Tonon,G. and Zuffi,G.
Recombinant bacterial strains for the production of natural nucleosides and modified analogues thereof
Patent: JP 2002533126-A 5 08-OCT-2002;
NORPHARMA SPA
COMMENT OS Artificial Sequence
PN JP 2002533126-A/5
PD 08-OCT-2002
PF 23-DEC-1999 JP 2000591198
PR 23-DEC-1998 IT MI 98A002792
PI GIUSEPPINA BESTETTI,SIMONA CALI,DANIELA GHISOTTI,GAETANO PI ORSINI,
PI GIANCARLO TONON,GABRIELE ZUFFI
PC C12N15/09,C12N1/21,C12N9/10//C12P19/38,C12P19/40,C12N15/00 CC Description of Artificial Sequence: Plasmid
CC udp Location/Qualifiers
FH Key Location/Qualifiers
FT Gene (243)..(1021)
FT Gene (1037)..(1766).
FEATURES source 1..4189
/organism="synthetic construct"
/mol_type="genomic DNA"
/db_xref="taxon:32630"
ORIGIN
Query Match 99.6%; Score 716.8; DB 6; Length 4189;
Best Local Similarity 99.7%; Pred. No. 1.8e-181;
Matches 718; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Qy 1 ATGGCTACCCACACATTAATGCAGAAATGGCGGATTCGCTGACGTAGTTTGGATGCCA 60
Db 1046 ATGGCTACCCACACATTAATGCAGAAATGGCGGATTCGCTGACGTAGTTTGGATGCCA 1105
Qy 61 GGGGACCCGCTGCGTGGGAGGTATATGCTCAAACTTCTTGAAGATGCCCGTGAAGTG 120
Db 1106 GGGGACCCGCTGCGTGGGAGGTATATGCTCAAACTTCTTGAAGATGCCCGTGAAGTG 1165
Qy 121 AACAACTGTCGCGGTATGCTGGGCTTCAACCGGTACTTACAAAGGCGCAAAATTTCCGTA 180
Db 1166 AACAACTGTCGCGGTATGCTGGGCTTCAACCGGTACTTACAAAGGCGCAAAATTTCCGTA 1225
Qy 181 ATGGGTACCGGTGTGGGTATCCCGTCTGCTGCATCTACACAAAGAACTGATCACCAGT 240
Db 1226 ATGGGTACCGGTATGGGTATCCCGTCTGCTGCATCTACACAAAGAACTGATCACCAGT 1285
Qy 241 TTCCGCGTGAAGAAATATTCGCGGTGGTTCCTGTCGCGGAGTTCTGCGCGCAGTAAAA 300
Db 1286 TTCCGCGTGAAGAAATATTCGCGGTGGTTCCTGTCGCGGAGTTCTGCGCGCAGTAAAA 1345
Qy 301 CTGCGCGACGTCGTTATCGGTATGGGTGCTGCACCGATTCCTCAAAAGTTAAACCGATCCGT 360
Db 1346 CTGCGCGACGTCGTTATCGGTATGGGTGCTGCACCGATTCCTCAAAAGTTAAACCGATCCGT 1405
Qy 361 TTTAAAGACCATGACTTTTGGCGGTATCGCTGACTTCGACATGCTGCGTGAAGAGTAGAT 420
Db 1406 TTTAAAGACCATGACTTTTGGCGGTATCGCTGACTTCGACATGCTGCGTGAAGAGTAGAT 1465
Qy 421 GCAGCTAAAGCAGTGGGTATGATGCTGCGGTGGGTAACTGTTCTCGCTGACCTGTTTC 480
Db 1466 GCAGCTAAAGCAGTGGGTATGATGCTGCGGTGGGTAACTGTTCTCGCTGACCTGTTTC 1525
Qy 481 TACTCTCCGGACGGCGAAATGTTTCGAGCTGATGGAATAATACGGCATTTCTCGGGGTGGAA 540
Db 1526 TACTCTCCGGACGGCGAAATGTTTCGAGCTGATGGAATAATACGGCATTTCTCGGGGTGGAA 1585

```
Qy 541 ATGGAAGCGGTGTATCTAGCGCTCGCTGCAGAAATTTGCGCGAAGCCCTGACCATC 600
    |||||||
Db 1586 ATGGAAGCGGTGTATCTAGCGCTCGCTGCAGAAATTTGCGCGAAGCCCTGACCATC 1645
    |||||||
Qy 601 TGCACCGTATCTGACCAATCCGCACTACAGAGCAGACCACTGCGCTGAGCGTCAGACT 660
    |||||||
Db 1646 TGCACCGTATCTGACCAATCCGCACTACAGAGCAGACCACTGCGCTGAGCGTCAGACT 1705
    |||||||
Qy 661 ACCTTCAACGACATGATCAAAATCGCACTCGAATCCGTTCTGTGGCGGATAAAGAGTAA 720
    |||||||
Db 1706 ACCTTCAACGACATGATCAAAATCGCACTCGAATCCGTTCTGTGGCGGATAAAGAGTAA 1765
    |||||||

RESULT 15
AX027813
LOCUS AX027813 4189 bp DNA linear PAT 16-SEP-2000
DEFINITION Sequence 5 from Patent WO0039307.
ACCESSION AX027813
VERSION AX027813.1 GI:10188657
KEYWORDS synthetic construct
SOURCE synthetic construct
ORGANISM other sequences; artificial sequences.
REFERENCE
1 Bestetti,G., Cali,S., Orsini,G., Tonon,G., Zuffi,G. and Ghisotti,D.
  Recombinant bacterial strains for the production of natural
  nucleosides and modified analogues thereof
  Patent: WO 0039307-A 5 06-JUL-2000;
  BESTETTI GIUSEPPINA (IT) ; CALI SIMONA (IT) ; NORPHARMA SPA (IT) ;
  ORSINI GAETANO (IT) ; TONON GIANCARLO (IT) ; ZUFFI GABRIELE (IT) ;
  GHISOTTI DANIELA (IT)
FEATURES
    source
    location/Qualifiers
    1..4189
    /organism="synthetic construct"
    /mol_type="unassigned DNA"
    /db_xref="taxon:32630"
    /note="Plasmid"

ORIGIN

Query Match 99.6%; Score 716.8; DB 6; Length 4189;
Best Local Similarity 99.7%; Pred. No. 1.8e-181;
Matches 718; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 ATGGCTACCCACACATTAATGAGAAATGGCGATTTCGCTGACGTAGTTTGTATGCCA 60
    |||||||
Db 1046 ATGGCTACCCACACATTAATGAGAAATGGCGATTTCGCTGACGTAGTTTGTATGCCA 1105
    |||||||
Qy 61 GGGACCCGCTGCGTGCAGATATATTGCTCAAACTTTCTTGAAGATGCCCGTGAAGTG 120
    |||||||
Db 1106 GGGACCCGCTGCGTGCAGATATATTGCTCAAACTTTCTTGAAGATGCCCGTGAAGTG 1165
    |||||||
Qy 121 AACAACTTCCGCGTATGCTGGGCTTCACCGGTACTTACAAAGCCGCCAAAATTTCCGTA 180
    |||||||
Db 1166 AACAACTTCCGCGTATGCTGGGCTTCACCGGTACTTACAAAGCCGCCAAAATTTCCGTA 1225
    |||||||
Qy 181 ATGGGTACCGGTGTGGTATCCGCTCCGTCCTGTCCTGTCGCGCAGTTCTGCGCGACGTAAATA 240
    |||||||
Db 1226 ATGGGTACCGGTATGGGTATCCCGTCTCTGCTCCATCTACACCAAGAACTGATCACCGAT 1285
    |||||||
Qy 241 TTGCGCGTGAAGAAAATTTATCCGCGTGGGTTCCTGTGGCGCAGTTCTGCGCGACGTAAATA 300
    |||||||
Db 1286 TTGCGCGTGAAGAAAATTTATCCGCGTGGGTTCCTGTGGCGCAGTTCTGCGCGACGTAAATA 1345
    |||||||
Qy 301 CTGCGCGACGTGCGTATCGGTA TGGGTGCTGCAACCGATTCCAAAGTTAAACCGATCCGT 360
    |||||||
Db 1346 CTGCGCGACGTGCGTATCGGTA TGGGTGCTGCAACCGATTCCAAAGTTAAACCGATCCGT 1405
    |||||||
Qy 361 TTTTAAACACCATGATTTGGCGTATCGCTGACTTCCGACATGGTGCCTAACGCGATAGAT 420
    |||||||
Db 1406 TTTTAAACACCATGATTTGGCGTATCGCTGACTTCCGACATGGTGCCTAACGCGATAGAT 1465
    |||||||
Qy 421 GCAGCTAAAGCAGCTGGGTATTGATGCTCGGCTGGGTAACCTGTTCTCCGCTGACCTGTTTC 480
    |||||||
```

```
Db 1466 GCAGCTAAAGCAGCTGGGTATTGATGCTCGCGTGGGTAACCTGTTCTCCGCTGACCTGTTTC 1525
    |||||||
Qy 481 TACTCTCCGACCGCGGAAATGTTTCGACGTGATCGGAAAATACCGCATTTCTCGGCGTGGAA 540
    |||||||
Db 1526 TACTCTCCGACCGCGGAAATGTTTCGACGTGATCGGAAAATACCGCATTTCTCGGCGTGGAA 1585
    |||||||
Qy 541 ATGGAAGCGGTGTATCTAGCGCTCGCTGCAGAAATTTGCGCGGAAAGCCCTGACCATC 600
    |||||||
Db 1586 ATGGAAGCGGTGTATCTAGCGCTCGCTGCAGAAATTTGCGCGGAAAGCCCTGACCATC 1645
    |||||||
Qy 601 TGCAACCGTATCTGACCAATCCGCACTACAGAGCAGACCACTGCGCTGAGCGTCAGACT 660
    |||||||
Db 1646 TGCAACCGTATCTGACCAATCCGCACTACAGAGCAGACCACTGCGCTGAGCGTCAGACT 1705
    |||||||
Qy 661 ACCTTCAACGACATGATCAAAATCGCACTCGAATCCGTTCTGTGGCGGATAAAGAGTAA 720
    |||||||
Db 1706 ACCTTCAACGACATGATCAAAATCGCACTCGAATCCGTTCTGTGGCGGATAAAGAGTAA 1765
    |||||||
```

Search completed: July 22, 2005, 21:18:03
Job time : 5511 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 21, 2005, 23:14:29 ; Search time 720 Seconds
(without alignments)

5919.741 Million cell updates/sec

Title: US-10-035-300A-1

Perfect score: 720

Sequence: 1 atggctaccacacattaa.....tgctggcgataaagagtaa 720

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N_Geneseq_16Dec04:*

1: Geneseqn1980s:*

2: Geneseqn1990s:*

3: Geneseqn2000s:*

4: Geneseqn2001as:*

5: Geneseqn2001bs:*

6: Geneseqn2002as:*

7: Geneseqn2002bs:*

8: Geneseqn2003as:*

9: Geneseqn2003bs:*

10: Geneseqn2003cs:*

11: Geneseqn2003ds:*

12: Geneseqn2004as:*

13: Geneseqn2004bs:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	720	100.0	720	8	AAD56237 Escherich
2	716.8	99.6	720	5	Aaf55440 Nucleotid
3	716.8	99.6	720	8	AAD54112 Escherich
4	716.8	99.6	720	12	Adl27841 E coli de
5	716.8	99.6	752	6	Ab151595 Purine nu
6	716.8	99.6	3383	3	Aaa51632 Plasmid p
7	716.8	99.6	4189	3	Aaa51634 Plasmid p
8	716.8	99.6	5013	9	ACA63355 E. coli D
9	716.8	99.6	5013	10	AAD59423 E.coli pu
10	716.8	99.6	5013	10	AdE86127 E. coli p
11	716.8	99.6	5495	3	AAA51633
12	716.8	99.6	6301	3	AAA51635
13	715.2	99.3	720	8	AAD56238
14	713.6	99.1	720	5	AAF55446
15	639.8	88.9	3066	3	AAA51641
16	638.8	88.7	5303	3	AAA51636
17	636.8	88.4	3165	3	AAA51642
18	636.8	88.4	3979	3	AAA51643
19	636.8	88.4	6117	3	AAA51644
20	632.8	87.9	6373	3	AAA51639

21	628.8	87.3	6343	3	AAA51638
22	570	79.2	2807	5	AAS86794
23	558.8	77.6	1112	5	AAS89111
24	558.4	77.6	732	11	ACH95836
25	533.6	74.1	1893	5	AAS88507
26	449.2	62.4	110000	10	ACF65385 0
c 27	449.2	62.4	110000	10	ACF67367_29
c 28	448.8	62.3	717	10	ACF69888
29	438.4	60.9	783	5	AAS86792
30	420.8	58.4	795	10	ADF02086
31	403.6	56.1	6701	13	ADT05517
32	403.6	56.1	85814	13	ADT05644
c 33	402.6	55.9	110000	2	AAT42063_05
34	330	45.8	5890	3	AAA51637
c 35	243.4	33.8	110000	6	ABA92787_5
36	212.2	29.5	110000	6	ABA90521_09
37	210.8	29.3	708	8	ACF73435
38	210.8	29.3	5301	2	AAV74754
39	206.6	28.7	2040	2	AAT77102
40	205.8	28.6	2678	5	AAS89114
41	205.4	28.5	899	13	ADRI2667
42	204.2	28.4	708	10	ABX06511
43	204.2	28.4	711	2	AAV72439
44	204.2	28.4	711	2	AAV72438
45	204.2	28.4	714	13	ADR93332

ALIGNMENTS

RESULT 1

AAD56237

ID AAD56237 standard; DNA; 720 BP.

AC AAD56237;

XX

DT 07-AUG-2003 (first entry)

DE

OS Escherichia coli mutant PNP enzyme (M65V) encoding DNA #SEQ ID 1.

XX

KW Purine nucleoside phosphorylase; mutant; tumour; PNP; enzyme; cancer;
KW leukaemia; glioblastoma; lymphoma; melanoma; carcinoma; gene therapy;
KW gene; ds.

OS Escherichia coli.

XX Synthetic.

FT Key Location/Qualifiers

FT CDS 1..720

FT /product= "Mutant PNP enzyme"

XX

PN WO2003035012-A2.

XX

PD 01-MAY-2003.

XX

PF 28-OCT-2002; 2002WO-US034626.

XX

PR 26-OCT-2001; 2001US-00035300.

XX

PA (UABR-) UAB RES FOUND.

PA (SOUR) SOUTHERN RES INST.

PA (CORR) CORNELL RES FOUND INC.

XX

PI Ealick SE, Parker WB, Seerist JA, Sorscher EJ;

XX

DR WPI; 2003-421350/39.

DR P-ESDB; AAE37217.

XX

PT Novel purified mutant purine cleaving enzyme having different biological
PT activity than a wild-type purine cleaving enzyme, useful for treating a
PT pathological condition characterized by abnormal cell growth.

XX

Db 121 AACAAAGTTCGGGTATGCTGGGCTTACCGGTACTTACAAAGCGCCGAAATTTCCGTA 180
Qy 181 ATGGTCAAGCTGTTGGTATCCCGTCCCTGCTCATACACCAAGAACTGATCACCGAT 240
Db 181 ATGGTCAAGCTGTTGGTATCCCGTCCCTGCTCATACACCAAGAACTGATCACCGAT 240
Qy 241 TTCGGGTGAAGAAATTTATCCCGTGGGTCTGCTGGGCGAGTTCTGCGGCACTGAAA 300
Db 241 TTCGGGTGAAGAAATTTATCCCGTGGGTCTGCTGGGCGAGTTCTGCGGCACTGAAA 300
Qy 301 CTGGCGAGCTGTTATCCGATGATGGTCTGCTGACCGGATTCGAAATTTACCGCATCCGT 360
Db 301 CTGGCGAGCTGTTATCCGATGATGGTCTGCTGACCGGATTCGAAATTTACCGCATCCGT 360
Qy 361 TTTAAGACCATGACTTTTCCGCTATCGCTGACTTTCGACATGGTGGTAAAGCAGTAGAT 420
Db 361 TTTAAGACCATGACTTTTCCGCTATCGCTGACTTTCGACATGGTGGTAAAGCAGTAGAT 420
Qy 421 GCAGCTAAGCACTGGGTATTTGATGCTCGGTGGGTAAACCTGTTCTCGCTGACCTGTT 480
Db 421 GCAGCTAAGCACTGGGTATTTGATGCTCGGTGGGTAAACCTGTTCTCGCTGACCTGTT 480
Qy 481 TACTCTCCGAGCGGCAATGTTGACCTGATGGAAATACGGCATTTCTGGCGTGGAA 540
Db 481 TACTCTCCGAGCGGCAATGTTGACCTGATGGAAATACGGCATTTCTGGCGTGGAA 540
Qy 541 ATGGAAGCGGTGGTATCTACGGCTCGCTGAGAAATTTGGCGGAAAGCCCTGACCATC 600
Db 541 ATGGAAGCGGTGGTATCTACGGCTCGCTGAGAAATTTGGCGGAAAGCCCTGACCATC 600
Qy 601 TGCACCGTATGACCATCGCACTCACGAGCAGACCACTGCGGCTGAGCGTCAGACT 660
Db 601 TGCACCGTATGACCATCGCACTCACGAGCAGACCACTGCGGCTGAGCGTCAGACT 660
Qy 661 ACCTTCACGACATGATCAAAATCGCACTGGAATCCGTTCTGCTGGCGGATAAAGAGTAA 720
Db 661 ACCTTCACGACATGATCAAAATCGCACTGGAATCCGTTCTGCTGGCGGATAAAGAGTAA 720

RESULT 3

AAD54112 standard; DNA; 720 BP.

AAD54112;
AC

17-JUN-2003 (first entry)

Escherichia coli cytosine nucleoside phosphorylase DNA.

Cytosine nucleoside phosphorylase; cytosine nucleoside compound;

pharmaceutical; enzyme; gene; ds.

Escherichia coli.

Key Location/Qualifiers

CDS 1..720

/*tag= a

/product= "E. coli cytosine nucleoside phosphorylase"

EPI254959-A2.

06-NOV-2002.

01-MAY-2002; 2002EP-00253075.

01-MAY-2001; 2001JP-00134352.

(MITA) MITSUI CHEM INC.

Araki T, Ikeda I, Matoishi K, Abe R, Oikawa T, Matsuba Y;

Ishibashi H, Nagahara K, Fukui Y;

WPI; 2003-158787/16.

RESULT 4

ADL27841

ID ADL27841 standard; DNA; 720 BP.

DR P-PSDB; AAE35397.

XX Producing cytosine nucleoside compound by reacting sugar phosphate and
PT cytosine or cytosine derivative in the presence of cytosine nucleoside
PT phosphorylase.

XX Disclosure; Page 17-18; 32pp; English.

XX The invention relates to a method for producing cytosine nucleoside
CC compound by reacting sugar phosphate and cytosine or cytosine derivative
CC in the presence of cytosine nucleoside phosphorylase. The method is used
CC to produce cytosine nucleoside compounds which are intermediates for
CC pharmaceuticals. The present sequence is Escherichia coli cytosine
CC nucleoside phosphorylase DNA

SQ Sequence 720 BP; 172 A; 190 C; 187 G; 171 T; 0 U; 0 Other;

Query Match 99.6%; Score 716.8; DB 8; Length 720;

Best Local Similarity 99.7%; Pred. No. 9.6e-217;

Matches 718; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 ATGGCTACCCACACATTAATGCGAAATGGCGGATTTTCGCTGACGTAGTTTGGATGCA 60

Db 1 ATGGCTACCCACACATTAATGCGAAATGGCGGATTTTCGCTGACGTAGTTTGGATGCA 60

Qy 61 GCGGACCGCTGCGTGGCAAGTATATTGCTGAAACTTTTCTTGAAGATGCCCGTGAAGTG 120

Db 61 GCGGACCGCTGCGTGGCAAGTATATTGCTGAAACTTTTCTTGAAGATGCCCGTGAAGTG 120

Qy 121 AACAAAGTTCGGGTATGCTGGGCTTACCGGTACTTACAAAGCGCGGAAATTTCCGTA 180

Db 121 AACAAAGTTCGGGTATGCTGGGCTTACCGGTACTTACAAAGCGCGGAAATTTCCGTA 180

Qy 181 ATGGGTCAAGCTGTTGGTATCCCGTCTGCTCATACCAAGAACTGATCACCGAT 240

Db 181 ATGGGTCAAGCTGTTGGTATCCCGTCTGCTCATACCAAGAACTGATCACCGAT 240

Qy 241 TTCGGCGTGAAGAAATTTATCCGGTGGGTTCTGTGGCGAGTTCTCGCGCACGTAATA 300

Db 241 TTCGGCGTGAAGAAATTTATCCGGTGGGTTCTGTGGCGAGTTCTCGCGCACGTAATA 300

Qy 301 CTGGCGAGCTGTTATCGGTATGGGTGCTGACCGGATTCGAAAGTTAACCGCATCGT 360

Db 301 CTGGCGAGCTGTTATCGGTATGGGTGCTGACCGGATTCGAAAGTTAACCGCATCGT 360

Qy 361 TTTAAGACCATGACTTTTGGCTATCGCTGACTTTCGACATGGTGGTAAAGCAGTAGAT 420

Db 361 TTTAAGACCATGACTTTTGGCTATCGCTGACTTTCGACATGGTGGTAAAGCAGTAGAT 420

Qy 421 GCAGCTAAGCACTGGGTATTGATGCTCGGTGGGTAAACCTGTTCTCGCTGACCTGTT 480

Db 421 GCAGCTAAGCACTGGGTATTGATGCTCGGTGGGTAAACCTGTTCTCGCTGACCTGTT 480

Qy 481 TACTCTCCGAGCGGCAATGTTGACGTGATGAAATACGGCATTTCTGGCGTGGAA 540

Db 481 TACTCTCCGAGCGGCAATGTTGACGTGATGAAATACGGCATTTCTGGCGTGGAA 540

Qy 541 ATGGAAGCGGTGGTATCTACGGGTGCTGCTGAGAAATTTGGCGGAAAGCCCTGACCATC 600

Db 541 ATGGAAGCGGTGGTATCTACGGGTGCTGCTGAGAAATTTGGCGGAAAGCCCTGACCATC 600

Qy 601 TGCACCGTATCTGACCATCATCCGACCTCACGAGCAGACCACTGCGGCTGAGCGTCA 660

Db 601 TGCACCGTATCTGACCATCATCCGACCTCACGAGCAGACCACTGCGGCTGAGCGTCA 660

Qy 661 ACCTTCACGACATGATCAAAATCGCACTGGAATCCGTTCTGCTGGCGGATAAAGAGTAA 720

Db 661 ACCTTCACGACATGATCAAAATCGCACTGGAATCCGTTCTGCTGGCGGATAAAGAGTAA 720

XX ADL27841;
AC 20-MAY-2004 (first entry)
DT E coli deod coding sequence.
XX ds; gene; selection method; yield; pathogen resistance;
KW nutritional quality; stress resistance; plant; transgenic.
XX Escherichia coli.
OS WO2004013333-A2.
XX 12-FEB-2004.
XX 18-JUL-2003; 2003WO-EP007877.
XX 26-JUL-2002; 2002DE-01034287.
XX (BADI) BASF PLANT SCI GMBH.
XX Kock M, Frank M, Badur R;
XX WPI; 2004-157134/15.
XX P-PSDB; ADL27842.
XX Method for transforming plant cells, useful for preparing transgenic
PT plants for e.g. food or production of chemicals, with selection based on
PT suppressing toxic effects of a marker protein.
XX Disclosure; Page 139-140; 201pp; German.
XX The present invention relates to a method for preparing transformed plant
CC cells or organisms, which comprises first transfecting a population of
CC plant cells, already containing at least one marker protein that can
CC exert a direct or indirect toxic effect, with at least one nucleic acid
CC sequence together with at least one compound able to reduce expression,
CC amount, activity and/or function of the marker protein. Transfected cells
CC that contain the nucleic acid in their genome and, because of the
CC activity of the compound have a growth advantage relative to non-
CC transformed cells are selected, under conditions where the marker protein
CC exerts its toxic effect on non-transformed cells. The method is
CC especially used to produce transgenic plants, especially those having an
CC advantageous phenotype, e.g. better nutritional quality for humans or
CC animals, production of selected chemicals or pharmaceuticals, increased
CC resistance to pathogens or environmental stress or higher yields. The
CC present sequence is a coding sequence shown in the exemplification of the
XX invention.
SQ Sequence 720 BP; 172 A; 190 C; 187 G; 171 T; 0 U; 0 Other;
Query Match 99.6%; Score 716.8; DB 12; Length 720;
Best Local Similarity 99.7%; Pred. No. 9.6e-217;
Matches 718; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 ATGGCTACCCACACATTATGCGAAGATGGCGGATTCGCTGACGAGTTTTCGATGCCA 60
DB 1 ATGGCTACCCACACATTATGCGAAGATGGCGGATTCGCTGACGAGTTTTCGATGCCA 60
QY 61 GGCACCGCGTGCCTGGGAGTATATGCTGAAACTTTCCCTTGAAGATGCCCGTGAAGTG 120
DB 61 GGCACCGCGTGCCTGGGAGTATATGCTGAAACTTTCCCTTGAAGATGCCCGTGAAGTG 120
QY 121 AACACGTTTCGCGTATGCTGGGCTTCACCGGTACTTACAAAGCCCGCAAAATTCGCTGA 180
DB 121 AACACGTTTCGCGTATGCTGGGCTTCACCGGTACTTACAAAGCCCGCAAAATTCGCTGA 180
QY 181 ATGGGTACCGGTGTTGGTATCCGCTCTGCTCCATCTACACCAAGAACTGATCACCAGAT 240
DB 181 ATGGGTACCGGTATGGGTATCCCGTCTGCTCCATCTACACCAAGAACTGATCACCAGAT 240
QY 241 TTCGGCGTGAAGAAAAATTTATCCGCGTGGGTTCTGTGGCGGAGTTCTGCGCGACGTAAAA 300

DB 241 TTCGGCGTGAAGAAAAATTTATCCGCGTGGGTTCTGTGGCGAGTTCTGCGCGACGTAAAA 300
QY 301 CTGCGGAGCTGCTTATCGGTATGGGTGCTGCACCGATTCCAAAGTTAAACCGCATCCGT 360
DB 301 CTGCGGAGCTGCTTATCGGTATGGGTGCTGCACCGATTCCAAAGTTAAACCGCATCCGT 360
QY 361 TTTAAAGACCATGACTTTGCGCTATCGCTGACTTCGACATGCTGCGTAAACGAGTAGAT 420
DB 361 TTTAAAGACCATGACTTTGCGCTATCGCTGACTTCGACATGCTGCGTAAACGAGTAGAT 420
QY 421 GCAGCTAAAGCACTGGGTATGCTGCGTGGGTAAACCTGTTCTCCGCTGACCTGTTTC 480
DB 421 GCAGCTAAAGCACTGGGTATGCTGCGTGGGTAAACCTGTTCTCCGCTGACCTGTTTC 480
QY 481 TACTCTCCGACCGCGAATGTTTCGACGTGATGGAATAATACGCGATTCCTCGCGCTGGAA 540
DB 481 TACTCTCCGACCGCGAATGTTTCGACGTGATGGAATAATACGCGATTCCTCGCGCTGGAA 540
QY 541 ATGGAAGCGGCTGCTATCTACGCGCTGCTGCGAATTTGGCGCGAAGCCCTGACCATC 600
DB 541 ATGGAAGCGGCTGCTATCTACGCGCTGCTGCGAATTTGGCGCGAAGCCCTGACCATC 600
QY 601 TGCACCGTATCTGACCATCCGACTTCACGACGACACCACTCCGCGTGAAGCTCAGACT 660
DB 601 TGCACCGTATCTGACCATCCGACTTCACGACGACACCACTCCGCGTGAAGCTCAGACT 660
QY 661 ACCTTCAAGACATGATCAAAATCGCACCTGGAATTCGTTCTGCTGGCGGATAAAGAGTAA 720
DB 661 ACCTTCAAGACATGATCAAAATCGCACCTGGAATTCGTTCTGCTGGCGGATAAAGAGTAA 720
RESULT 5
ABL51595
ID ABL51595 standard; DNA; 752 BP.
AC ABL51595;
XX 04-JUL-2002 (first entry)
DT Purine nucleoside phosphorylase related DNA sequence.
XX Purine nucleoside phosphorylase; enzyme; molecular biology;
XX biotechnology; genetic engineering; gene; ds.
XX Escherichia coli.
OS Synthetic.
XX Key Location/Qualifiers
FH primer_bind 7..35
FT /*tag= a
FT /*note= "oligonucleotide binding site"
FT primer_bind complement(709..742)
FT /*tag= b
FT /*note= "oligonucleotide binding site"
XX RU2179188-C2.
XX 10-FEB-2002.
XX 03-MAR-2000; 2000RU-00105214.
XX 03-MAR-2000; 2000RU-00105214.
XX (ASBI=) AS RUSSIA BIO-ORGANIC CHEM INST.
XX Bapov RS, Gurevich AI, Miroshnikov AI, Chuvikovskii DV;
XX WPI; 2002-265551/31.
XX Production of recombinant purine nucleoside phosphorylase, recombinant
PT plasmid DNA pERPUPH01 and strain Escherichia coli BL21 (DE3) perpuh01 for
PT its realization.

XX Disclosure; Col 9-10; Opp; Russian.

CC The present invention describes a recombinant plasmid DNA PERPUPH01 which encodes an amino acid sequence of an Escherichia coli purine nucleoside phosphorylase. The recombinant plasmid DNA PERPUPH01 consists of: a NcoI/EcoRI-fragment of plasmid pET23d DNA containing a promoter and a transcription terminator of T7 RNA polymerase, a translation enhancer of gene 10 of phage T7, a beta-lactamase gene and a NcoI/EcoRI-fragment of a DNA containing the Escherichia coli purine nucleoside phosphorylase gene sequence adapted to these sites. The recombinant plasmid DNA PERPUPH01 can be used in molecular biology, biotechnology and genetic engineering.

CC The present sequence represents a nucleotide sequence given in the exemplification of the present invention

XX

SQ Sequence 752 BP; 184 A; 194 C; 195 G; 179 T; 0 U; 0 Other;

Query Match 99.6%; Score 716.8; DB 6; Length 752;

Best Local Similarity 99.7%; Pred. No. 9.8e-217;

Matches 718; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ATGGCTACCCACACATTAATGAGAAATGGCGGATTTGCGTACGATAGTTTGGATGCCA 60

Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 13 ATGGCTACCCACACATTAATGAGAAATGGCGGATTTGCGTACGATAGTTTGGATGCCA 72

Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 61 GCGGACCCGCTGCGTGCAGAGTATATGCTGAAACTTTCTTTGAAAGATGCCCGTGAAGTG 120

Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 73 GCGGACCCGCTGCGTGCAGAGTATATGCTGAAACTTTCTTTGAAAGATGCCCGTGAAGTG 132

Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 121 AACACGTTCCGGGTATGCTGGGCTTACCGGTACTTACAAAGCGCGCAAAATTTCCGTA 180

Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 133 AACACGTTCCGGGTATGCTGGGCTTACCGGTACTTACAAAGCGCGCAAAATTTCCGTA 192

Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 181 ATGGGTACCGGTGTTGGTATCCCGTCTGCTCATCTACACCAAGAACTGATCACCGAT 240

Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 193 ATGGGTACCGGTATGGGTATCCCGTCTGCTCATCTACACCAAGAACTGATCACCGAT 252

Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 241 TTCGCGGTGAAGAAATATCCGCTGGGTTCCTGTGGCGCATTTCTCCGCGACGTAAAA 300

Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 253 TTCGCGGTGAAGAAATATCCGCTGGGTTCCTGTGGCGCATTTCTCCGCGACGTAAAA 312

Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 301 CTGCGCGAGCGTATATCGGTATGGGTGCTGCACCGATTCGAAATTAACGGCATCCGT 360

Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 313 CTGCGCGAGCGTATATCGGTATGGGTGCTGCACCGATTCGAAATTAACGGCATCCGT 372

Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 361 TTAAAGACCATGACTTTGGCGCTATCGTCACTTCGACATGTCGTAACGCATAGAT 420

Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 373 TTAAAGACCATGACTTTGGCGCTATCGTCACTTCGACATGTCGTAACGCATAGAT 432

Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 421 GCAGCTAAAGCACCTGGGTATTGATGCTCGCGTGGGTAACTGTTCTCCGCTGACCTGTT 480

Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 433 GCAGCTAAAGCACCTGGGTATTGATGCTCGCGTGGGTAACTGTTCTCCGCTGACCTGTT 492

Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 481 TACTCTCGGACGGCGAAATGTTTCAGCTGATGGAATAATACGGCATTTCTCGCGGTGAA 540

Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 493 TACTCTCGGACGGCGAAATGTTTCAGCTGATGGAATAATACGGCATTTCTCGCGGTGAA 552

Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 541 ATGGAAGCGGCTGATCTACGGGCTGCTGCAGAAATTTGGCGAAGAGCCCTGACCATC 600

Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 553 ATGGAAGCGGCTGATCTACGGGCTGCTGCAGAAATTTGGCGAAGAGCCCTGACCATC 612

Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 601 TGCACCGTATCTGACCATCGCATCTCACGAGCAGACCACTCGCGCTGAGCGTCCAGACT 660

Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 613 TGCACCGTATCTGACCATCGCATCTCACGAGCAGACCACTCGCGCTGAGCGTCCAGACT 672

Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 661 ACCTTCAACGACATGATCAAAATCGCATCGGATCCGTTCTGCTGGCGGATTAAGAGTAA 720

Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 673 ACCTTCAACGACATGATCAAAATCGCATCGGATCCGTTCTGCTGGCGGATTAAGAGTAA 732

Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 6
AAA51632
ID AAA51632 standard; DNA; 3383 BP.

Query Match 99.6%; Score 716.8; DB 3; Length 3383;
Best Local Similarity 99.7%; Pred. No. 2e-216;
Matches 718; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

XX AAA51632;

XX 31-OCT-2000 (first entry)

DE Plasmid pGM678 containing E. coli deoD gene.

XX Uridine phosphorylase; udp; purine nucleoside phosphorylase; PNP; deoD;
Genetically modified organism; catalyst; transglycosylation;
nucleoside analogue; alpha-pentose-1-phosphate sugar; phosphorolysis;
anti-viral; anti-tumour; mesophilic bacterium; ss.

XX Escherichia coli.
OS Synthetic.

XX

FH Key Location/Qualifiers
FT misc_RNA 1..230
FT /tag= a
FT /label= pUC18_sequence
FT CDS 216..952
FT /tag= b
FT /product= "lacZ-deoD_fused_gene"
FT misc_RNA 961..3383
FT /tag= c
FT /label= pUC18_sequence

XX WO200039307-A2.

XX 06-JUL-2000.

XX 23-DEC-1999; 99WO-EP010416.

XX 23-DEC-1998; 98IT-MI002792.

XX (NORP-) NORPHARMA SPA.

XX Beatetti G, Cali' S, Ghisotti D, Orsini G, Tonon G, Zuffi G;
WPI; 2000-452402/39.

XX Recombinant expression vectors used to transform cells for the production of e.g. nucleosides encode uridine phosphorylase and/or purine nucleoside phosphorylase.

PS Claim 14; Page 47-48; 72pp; English.

XX Host cells genetically modified to express uridine phosphorylase (Udp) and purine nucleoside phosphorylase (PNP) or their corresponding crude or purified extracts, either separately or in combination are used as catalysts of transglycosylation reactions between a donor nucleoside and an acceptor base, for preparing nucleoside analogues containing heterocyclic systems with purine and/or pyrimidine bases substituted by one or more nitrogen atoms, for preparing alpha-pentose-1-phosphate sugars by phosphorolysis reactions and for producing nucleosides and modified nucleoside analogues (all claimed). The modified or natural nucleosides are used directly or as intermediates in the preparation of drugs with anti-viral and anti-tumour activity and for preparing oligonucleotides for therapeutic or diagnostic use. The recombinant host cells are produced using recombinant plasmid expression vectors comprising at least one gene sequence of a mesophilic bacterium coding for a polypeptide with UDP activity and/or enzyme PNP activity and at least one gene sequence coding for antibiotic resistance. In particular, the E. coli udp and deoD genes are used. Recombinant strains produced using the vectors express polypeptides with enzyme Udp and PNP activity in large amounts, e.g. 340-1040 times higher Udp activity and 120-200 times higher PNP activity than non-transformed corresponding wild type strains

XX Sequence 3383 BP; 826 A; 868 C; 855 G; 834 T; 0 U; 0 Other;

Qy	1	ATGGCTACCCACACATTAATGCAGAAATGGCGATTTCGCTGACGTAGTTTTTGATGCCA	60
Ds	240	ATGGCTACCCACACATTAATGCAGAAATGGCGATTTCGCTGACGTAGTTTTTGATGCCA	299
Qy	61	GGCGACCGCTGCGTGGCGAAGTATATTGCTGAAACTTTTCCCTTGAAGATGCCCGTGAAGTG	120
Ds	300	GGCGACCGCTGCGTGGCGAAGTATATTGCTGAAACTTTTCCCTTGAAGATGCCCGTGAAGTG	359
Qy	121	AACAAAGTTTCGCGGTATGCTGGGCTTACCGGTACTTACAAAGCCCGCAAAATTTCCGTA	180
Ds	360	AACAAAGTTTCGCGGTATGCTGGGCTTACCGGTACTTACAAAGCCCGCAAAATTTCCGTA	419
Qy	181	ATGGGTACCGGTGTTGGGTATCCCGTCTGCTCCATCTACACCAAGAACTGATCACCGAT	240
Ds	420	ATGGGTACCGGTGTTGGGTATCCCGTCTGCTCCATCTACACCAAGAACTGATCACCGAT	479
Qy	241	TTCCGCGTGAAGAAATATATCCGGTGGGTTCTGTGGCGAGTTCTGCGCGACGTAAAA	300
Ds	480	TTCCGCGTGAAGAAATATATCCGGTGGGTTCTGTGGCGAGTTCTGCGCGACGTAAAA	539
Qy	301	CTGGCGACGTCGTATCGGTATGGGTGCTGCGACCGATTCCAAAGTTAACCGCATCGGT	360
Ds	540	CTGGCGACGTCGTATCGGTATGGGTGCTGCGACCGATTCCAAAGTTAACCGCATCGGT	599
Qy	361	TTTAAAGACCATGACTTTGCCGCTATCGCTGACTTCGACATGCGTAAAGCAGTAGAT	420
Ds	600	TTTAAAGACCATGACTTTGCCGCTATCGCTGACTTCGACATGCGTAAAGCAGTAGAT	659
Qy	421	GCAGCTAAAGCACCTGGGTATTGATGCTCGCGTGGGTAACTGTTCTCGCTGACCTGTT	480
Ds	660	GCAGCTAAAGCACCTGGGTATTGATGCTCGCGTGGGTAACTGTTCTCGCTGACCTGTT	719
Qy	481	TACTCTCGGACGGCGAAATGTTGACGTGATGAGAAATACGGCATTCGCGGTGGAA	540
Ds	720	TACTCTCGGACGGCGAAATGTTGACGTGATGAGAAATACGGCATTCGCGGTGGAA	779
Qy	541	ATGGAAGCGGCTGCTATCTACGGGTGCTGTCAGAAATTTGGCGGAAGCCCTGACCATC	600
Ds	780	ATGGAAGCGGCTGCTATCTACGGGTGCTGTCAGAAATTTGGCGGAAGCCCTGACCATC	839
Qy	601	TGCACCGTATCTGACCATCCGCACTCAGCAGCAGACCACTGCGCGTGAAGCTCAGACT	660
Ds	840	TGCACCGTATCTGACCATCCGCACTCAGCAGCAGACCACTGCGCGTGAAGCTCAGACT	899
Qy	661	ACCTTCAACGACATGATCAAAATCGCACTGGAATCCGTTCTGCTGGGCGATAAAGATAA	720
Ds	900	ACCTTCAACGACATGATCAAAATCGCACTGGAATCCGTTCTGCTGGGCGATAAAGATAA	959
RESULT 7			
AAAS1634	ID	AAAS1634	standard; DNA; 4189 BP.
XX	AC	AAAS1634;	
XX	DT	31-OCT-2000	(first entry)
XX	DE	Plasmid pGM712	containing udp and deoD genes.
KW	KW	Uridine phosphorylase; udp; purine nucleoside phosphorylase; PNP; deoD;	
KW	KW	Genetically modified organism; catalyst; transglycosylation; anti-vital;	
KW	KW	nucleoside analogue; alpha-pentose-1-phosphate sugar; phosphorolysis;	
XX	XX	anti-tumour; mesophilic bacterium; tetracycline resistance; ss.	
OS	OS	Escherichia coli.	
OS	OS	Synthetic.	
XX	XX	Chimeric.	
FH	key	Location/Qualifiers	
FT	FT	1..242	
FT	FT	misc_RNA	
FT	FT	/*tag= a	
FT	FT	/label= pUC18_sequence	

FT	misc_RNA	243..1021	
FT	FT	/*tag= b	
FT	FT	/label= udp_gene_sequence	
FT	FT	1022..1025	
FT	FT	/*tag= c	
FT	FT	/label= pUC18_sequence	
FT	FT	1026..1036	
FT	FT	/*tag= d	
FT	FT	/label= pBAD24_sequence	
FT	FT	1037..1166	
FT	FT	/*tag= e	
FT	FT	/label= deoD_gene_sequence	
FT	FT	1167..11792	
FT	FT	/*tag= f	
FT	FT	/label= pBAD24_sequence	
FT	FT	11793..4189	
FT	FT	/*tag= g	
FT	FT	/label= pUC18_sequence	
XX	XX	WO200039307-A2.	
PN	PN	06-JUL-2000.	
XX	XX	23-DEC-1999;	99WO-EP010416.
XX	XX	23-DEC-1998;	98IT-MI002792.
XX	XX	(NORP-) NORPHARMA SPA.	
XX	XX	Bestetti G, Cali' S, Ghisotti D, Orsini G, Tonon G, Zuffi G;	
XX	XX	WPI; 2000-452402/39.	
XX	XX	Recombinant expression vectors used to transform cells for the production	
PT	PT	of e.g. nucleosides encode uridine phosphorylase and/or purine nucleoside	
PT	PT	phosphorylase.	
XX	XX	Claim 14; Page 51-52; 72pp; English.	
XX	XX	Host cells genetically modified to express uridine phosphorylase (Udp)	
CC	CC	and purine nucleoside phosphorylase (PNP) or their corresponding crude or	
CC	CC	purified extracts, either separately or in combination are used as	
CC	CC	catalysts of transglycosylation reactions between a donor nucleoside and	
CC	CC	an acceptor base, for preparing nucleoside analogues containing	
CC	CC	heterocyclic systems with purine and/or pyrimidine bases substituted by	
CC	CC	one or more nitrogen atoms, for preparing alpha-pentose-1-phosphate	
CC	CC	sugars by phosphorylase reactions and for producing nucleosides and	
CC	CC	modified nucleoside analogues (all claimed). The modified or natural	
CC	CC	nucleosides are used directly or as intermediates in the preparation of	
CC	CC	drugs with anti-viral and anti-tumour activity and for preparing	
CC	CC	oligonucleotides for therapeutic or diagnostic use. The recombinant host	
CC	CC	cells are produced using recombinant plasmid expression vectors	
CC	CC	comprising at least one gene sequence of a mesophilic bacterium coding	
CC	CC	for a polypeptide with UDP activity and/or enzyme PNP activity and at	
CC	CC	least one gene sequence coding for antibiotic resistance. In particular,	
CC	CC	the E. coli udp and deoD genes are used. Recombinant strains produced	
CC	CC	using the vectors express polypeptides with enzyme UDP and PNP activity	
CC	CC	in large amounts, e.g. 340-1040 times higher UDP activity and 120-200	
CC	CC	times higher PNP activity than non-transformed corresponding wild type	
CC	CC	strains	
XX	XX	Sequence 4189 BP; 998 A; 1077 C; 1081 G; 1033 T; 0 U; 0 Other;	
Qy	Query Match	99.6%; Score 716.8; DB 3; Length 4189;	
Qy	Best Local Similarity	99.7%; Pred. No. 2.3e-216;	
Qy	Matches 718; Conservative	0; Mismatches 2; Indels 0; Gaps 0;	
Qy	1	ATGGCTACCCACACATTAATGCAGAAATGGCGATTTCGCTGACGTAGTTTTTGATGCCA	60
Ds	1046	ATGGCTACCCACACATTAATGCAGAAATGGCGATTTCGCTGACGTAGTTTTTGATGCCA	1105
Qy	61	GGCGACCGCTGCGTGGCGAAGTATATTGCTGAAACTTTTCCCTTGAAGATGCCCGTGAAGTG	120

Db 1106 GCGGACCGCTGCGTGGGAAAGTATATGCTGTAACCTTCTTGAAGATGCCCGTGAAGTG 1165
Qy 121 AACACGTTTCGGGTATGCTGGGCTTTCACCGGTACTTACAAAGCGCGAAAATTTCCGTA 180
Db 1166 AACACGTTTCGGGTATGCTGGGCTTTCACCGGTACTTACAAAGCGCGAAAATTTCCGTA 1225
Qy 181 ATGGGTACGGTGTGGTATCCCGTCTGCTGCTCATCTACCAAGAAAGTATACCGAT 240
Db 1226 ATGGGTACGGTATGGGTATCCCGTCTGCTCATCTACCAAGAAAGTATACCGAT 1285
Qy 241 TTCGCGGTGAAGAAATTTATCCGCTGGGTCTGCTGGCGGAGTTCTCGCGCAGCTAAAA 300
Db 1286 TTCGCGGTGAAGAAATTTATCCGCTGGGTCTGCTGGCGGAGTTCTCGCGCAGCTAAAA 1345
Qy 301 CTGCGCGACGCTGGTATCGGTATGGGTGCTGCTGACCGGATTCCAAAGTTAAACCGCATCCGT 360
Db 1346 CTGCGCGACGCTGGTATCGGTATGGGTGCTGCTGACCGGATTCCAAAGTTAAACCGCATCCGT 1405
Qy 361 TTTAAAGACATGACTTTGGCGGTATCGGTGACTTCGACATGGTGGTAAAGCGATAGAT 420
Db 1406 TTTAAAGACATGACTTTGGCGGTATCGGTGACTTCGACATGGTGGTAAAGCGATAGAT 1465
Qy 421 GCAGCTAAGCAGCTGGGTATGATGCTCGCGTGGGTACCTGTTCTCGCTGACCTGTTTC 480
Db 1466 GCAGCTAAGCAGCTGGGTATGATGCTCGCGTGGGTAACTGTTCTCGCTGACCTGTTTC 1525
Qy 481 TACTCTCCGCGACGCGGAAATTTCCAGCTGATGTAAGAAATACCGCATTTCTCGGCGTGAA 540
Db 1526 TACTCTCCGCGACGCGGAAATTTCCAGCTGATGTAAGAAATACCGCATTTCTCGGCGTGAA 1585
Qy 541 ATGGAAGCGGTGGTATCTACGGGTGCTGCTGAGAAATTTGGCGGAAAGCCCTGACCATC 600
Db 1586 ATGGAAGCGGTGGTATCTACGGGTGCTGCTGAGAAATTTGGCGGAAAGCCCTGACCATC 1645
Qy 601 TGCACCGTATCTGACCATCGGCTACGACGACGACGACGACGACGACGACGACGACGAC 660
Db 1646 TGCACCGTATCTGACCATCGGCTACGACGACGACGACGACGACGACGACGACGACGAC 1705
Qy 661 ACCTTCAACGACATGATCAAAATCGCACTCGCAATCCGTTCTGCTGGGCGGATAAAGAGTAA 720
Db 1706 ACCTTCAACGACATGATCAAAATCGCACTCGCAATCCGTTCTGCTGGGCGGATAAAGAGTAA 1765

RESULT 8

ACA63355
ID ACA63355 standard; DNA; 5013 BP.

XX ACA63355;

XX 29-AUG-2003 (first entry)

XX E. coli DNA encoding PNP inserted in plasmid pTRCNP.

DE Prokaryotic host cell; transformation; transfection; vector; PNP;
XX purine nucleoside phosphorylase; hydrolase; replicating cell;
KW nonreplicating cell; bystander cell; purine substrate; tumour cell;
KW virally infected cell; cancer therapy; cytostatic; mutant; pTRCNP; ds.

OS Escherichia coli; strain JM101.

OS Synthetic.

XX US6491905-B1.

XX 10-DEC-2002.

XX 30-OCT-1998; 98US-00183188.

XX 14-SEP-1993; 93US-00122321.

PR 23-AUG-1996; 96US-00702181.

PR 24-JUN-1997; 97US-00881772.

PR 31-OCT-1997; 97US-0064676P.

XX (UABR-) UAB RES FOUND.

PA (SOUR) SOUTHERN RES INST.

XX Sorscher EJ, Parker WB, Waud W, Gadi VK, Bennett LL;

XX MPI; 2003-502635/47.

XX Novel prokaryotic host cell for treating tumor cells and virally infected
PT cells transfected by a vector comprising a DNA sequence
PT encoding Escherichia coli derived purine nucleoside phosphorylase.

XX Claim 1; Fig 15; 45pp; English.

XX The present invention relates to a prokaryotic host cell transformed or
CC transfected by a vector comprising a DNA sequence encoding Escherichia
CC coli derived purine nucleoside phosphorylase (PNP) or hydrolase. The
CC vector is useful for killing targeted replicating or nonreplicating
CC mammalian cells, and bystander cells. The host cell is useful in
CC combination with a purine substrate for treating tumour and/or virally
CC infected cells. The vector is useful in cancer therapy. The present
CC sequence represents DNA encoding E.coli PNP inserted in plasmid pTRCNP

XX Sequence 5013 BP; 1231 A; 1258 C; 1325 G; 1199 T; 0 U; 0 Other;

Query Match 99.6%; Score 716.8; DB 9; Length 5013;

Best Local Similarity 99.7%; Pred. No. 2.5e-216;

Matches 718; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 ATGCTACCCACACATTAATGCAGAAATGGCGGATTTCCGCTGACGTAGTTTTCATGCCA 60

Db 413 ATGCTACCCACACATTAATGCAGAAATGGCGGATTTCCGCTGACGTAGTTTTCATGCCA 472

Qy 61 GCGACCGCGTCCGTGGGAGTATATTGCTGAAACTTTCTTGAAGATGCCCGTGAAGTG 120

Db 473 GCGACCGCGTCCGTGGGAGTATATTGCTGAAACTTTCTTGAAGATGCCCGTGAAGTG 532

Qy 121 AACACGTTTCGGGTATGCTGGGCTTCCCGGTACTTACAAAGCGCGAAAATTTCCGTA 180

Db 533 AACACGTTTCGGGTATGCTGGGCTTCCCGGTACTTACAAAGCGCGAAAATTTCCGTA 592

Qy 181 ATGGGTACCGGTGTTGGTATCCCGTCTGCTCATCTACCAAGAACTGATCACCAGT 240

Db 593 ATGGGTACCGGTGTTGGTATCCCGTCTGCTCATCTACCAAGAACTGATCACCAGT 652

Qy 241 TTCGGCGTGAAGAAAATTTATCCGCGTGGGTTCTGTGGCGCAGTTTCGCCGACGTAAAA 300

Db 653 TTCGGCGTGAAGAAAATTTATCCGCGTGGGTTCTGTGGCGCAGTTTCGCCGACGTAAAA 712

Qy 301 CTGCGGACGCTGTTATCGGTATGGGTGCTGACCGATTCCAAAGTTAACCGCATCCGT 360

Db 713 CTGCGGACGCTGTTATCGGTATGGGTGCTGACCGATTCCAAAGTTAACCGCATCCGT 772

Qy 361 TTTAAAGACCATGACTTTGCCGCTATCGCTGACTTTCGACATGGTGGTAAACGCGATAGAT 420

Db 773 TTTAAAGACCATGACTTTGCCGCTATCGCTGACTTTCGACATGGTGGTAAACGCGATAGAT 832

Qy 421 GCACCTAAAGCAGCTGGGTATTGATGCTCGCGTGGGTAACTGTTCTCGCGTACCGTGTTC 480

Db 833 GCACCTAAAGCAGCTGGGTATTGATGCTCGCGTGGGTAACTGTTCTCGCGTACCGTGTTC 892

Qy 481 TACTCTCGGACGCGGAAATTTTCGAGTGTATGAAATAACGGGATTTCTCGGCGTGAA 540

Db 893 TACTCTCGGACGCGGAAATTTTCGAGTGTATGAAATAACGGGATTTCTCGGCGTGAA 952

Qy 541 ATGGAAGCGGTGGTATCTACGGGTGCTGTCAGAAATTTGGCGGAAAGCCCTGACCATC 600

Db 953 ATGGAAGCGGTGGTATCTACGGGTGCTGTCAGAAATTTGGCGGAAAGCCCTGACCATC 1012

Qy 601 TGCAACCGTATCTGACCAATCCGACATCCGACGAGCAGACCACTGCGCGTGGCGTACAGT 660

Db 1013 TGCAACCGTATCTGACCAATCCGACATCCGACGAGCAGACCACTGCGCGTGGCGTACAGT 1072

Qy 661 ACCTTCAACGACATGATCAAAATCGCACTGGAATCCGTTCTGCTGGGCGATTAAGAGTAA 720

Db 661 ACCTTCAACGACATGATCAAAATCGCACTGGAATCCGTTCTGCTGGGCGATTAAGAGTAA 720

```
Db 1073 ACCTTCAACGACATGATCAAAATCGCACTGGAATCCGTTCTGTGGGCGATAAAGACTAA 1132

RESULT 9
AAD59423
ID AAD59423 standard; DNA; 5013 BP.
XX
AC AAD59423;
XX
DT 18-DEC-2003 (first entry)
XX
DE E.coli purine analogue nucleoside phosphorylase DNA.
XX
KW Cytostatic; virucide; purine cleavage enzyme; tumour; therapy; PNP;
KW drug screening; purine analogue nucleoside phosphorylase; ds.
XX
OS Escherichia coli.
XX
FN US2003077268-A1.
XX
PD 24-APR-2003.
XX
PF 18-JUL-2002; 2002US-00198034.
XX
PR 14-SEP-1993; 93US-00122321.
PR 23-AUG-1996; 96US-00702181.
PR 24-JUN-1997; 97US-00881772.
PR 31-OCT-1997; 97US-0064676P.
PR 30-OCT-1998; 98US-00183188.
XX
(SORS/) SORSCHER E J.
PA (PARK/) PARKER W B.
PA (WAUD/) WAUD W.
PA (BENN/) BENNETT L L.
XX
PI Sorscher EJ, Parker WB, Waud W, Gadi VK, Bennett LL;
XX
DR WPI; 2003-755064/71.
XX
XX
XX Killing (non-)replicating targeted mammalian cells and bystander cells
XX useful for treating tumor cells and/or virally infected cells comprises
XX delivering purine nucleotide phosphorylase or hydrolase cleavage enzyme
XX to cells.
XX
PS Disclosure; Page 20-23; 40pp; English.
XX
XX The present invention relates to a method of killing replicating or non-
XX replicating, targeted mammalian cells and bystander cells, comprising
XX delivering a purine cleavage enzyme to the targeted mammalian cells and
XX contacting the targeted cells with a purine cleavage enzyme substrate to
XX kill the targeted cells when cleaved by the enzyme. The purine cleavage
XX enzyme is used for killing replicating or non-replicating targeted
XX mammalian cells and bystander cells. They are used for treating tumour
XX cells and virally infected cells. The method may also be useful in drug
XX screening. This sequence represents Escherichia coli PNP (purine analogue
XX nucleoside phosphorylase) DNA
XX
SQ Sequence 5013 BP; 1231 A; 1258 C; 1325 G; 1199 T; 0 U; 0 Other;

Query Match 99.6%; Score 716.8; DB 10; Length 5013;
Beet Local Similarity 99.7%; Pred. No. 2.5e-216;
Matches 718; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 ATGGCTACCCACACATTAATGCGAAATGGCGCAATTTCGCTGACGTAGTTTTCATGCCA 60
Db 413 ATGGCTACCCACACATTAATGCGAAATGGCGCAATTTCGCTGACGTAGTTTTCATGCCA 472
Qy 61 GGCGACCCGCTGCGTGGCAATATATGCTCAAACTTTCCTTGAAGATCCCGTGAAGTG 120
Db 473 GGCGACCCGCTGCGTGGCAATATATGCTCAAACTTTCCTTGAAGATCCCGTGAAGTG 532
Qy 121 AACCAACGTTTCGCGGTATGCTGGGCTTCACCGGTACTTACAAAGGCCCGCAAAATTTCCGTA 180
```

```
Db 533 AACCAAGTTTCGCGGTATGCTGGGCTTTCACCGGTACTTACAAAGGCCCGCAAAATTTCCGTA 592
Qy 181 ATGGGTACCGGTGTTGGTATCCCGTCTCTCATCTACACCAAGAACTGATCCCGAT 240
Db 593 ATGGGTACCGGTATGGGTATCCCGTCTCTCATCTACACCAAGAACTGATCCCGAT 652
Qy 241 TTCGGGTGAAGAAAATTTATCCGCGTGGGTTCTGTGGGCGAGTTCTGCGCGACGTAAAA 300
Db 653 TTCGGGTGAAGAAAATTTATCCGCGTGGGTTCTGTGGGCGAGTTCTGCGCGACGTAAAA 712
Qy 301 CTGCGGACGTCGTTATCGGTATGGGTGCTGCACCGATTCCAAAGTTTAAACCGATCCGT 360
Db 713 CTGCGGACGTCGTTATCGGTATGGGTGCTGCACCGATTCCAAAGTTTAAACCGATCCGT 772
Qy 361 TTTAAAGACCATGACATTTGCCGCTATCGCTGATGCTGACATGCTGCGTAAACGCGTAGAT 420
Db 773 TTTAAAGACCATGACATTTGCCGCTATCGCTGATGCTGACATGCTGCGTAAACGCGTAGAT 832
Qy 421 GCAGCTAAAGCACTGGGTATTGATGCTCGCGTGGGTAAACCTGTTCTCCGCTGACCTGTTTC 480
Db 833 GCAGCTAAAGCACTGGGTATTGATGCTCGCGTGGGTAAACCTGTTCTCCGCTGACCTGTTTC 892
Qy 481 TACTCTCCGACCGCGGAAATGTTTCGACGTGATCGAATAAATAACGGCATTTCTCGCGGTGAA 540
Db 893 TACTCTCCGACCGCGGAAATGTTTCGACGTGATCGAATAAATAACGGCATTTCTCGCGGTGAA 952
Qy 541 ATGGAAGCGCTGTTATCTACGGCGTCTGTCAGAAATTTGGCGGGAAGCCCTGACCATC 600
Db 953 ATGGAAGCGCTGTTATCTACGGCGTCTGTCAGAAATTTGGCGGGAAGCCCTGACCATC 1012
Qy 601 TGCAACGTTATCTGACCACTCCGCACTCAGCAGCAGACCACTCCGCTGAGCGCTCAGACT 660
Db 1013 TGCAACGTTATCTGACCACTCCGCACTCAGCAGCAGACCACTCCGCTGAGCGCTCAGACT 1072
Qy 661 ACCTTCAACGACATGATCAAAATCGCACTGGAATCCGTTCTGTGGGCGATAAAGAGTAA 720
Db 1073 ACCTTCAACGACATGATCAAAATCGCACTGGAATCCGTTCTGTGGGCGATAAAGAGTAA 1132

RESULT 10
ADE86127
ID ADE86127 standard; DNA; 5013 BP.
XX
AC ADE86127;
XX
DT 29-JAN-2004 (first entry)
XX
DE E. coli purine nucleoside phosphorylase (PNP) gene.
XX
KW Tumour growth inhibition; purine analogue; stable transformation;
KW prokaryotic host cell; purine nucleoside phosphorylase; PNP;
KW purine nucleoside hydrolase; tumour cell; virally infected cell;
KW toxin gene therapy; cytostatic; virucide; gene; ds.
XX
OS Escherichia coli.
XX
FN US2003134819-A1.
XX
PD 17-JUL-2003.
XX
PF 18-JUL-2002; 2002US-00198228.
XX
PR 14-SEP-1993; 93US-00122321.
PR 23-AUG-1996; 96US-00702181.
PR 24-JUN-1997; 97US-00881772.
PR 31-OCT-1997; 97US-0064676P.
PR 30-OCT-1998; 98US-00183188.
XX
(SORS/) SORSCHER E J.
PA (PARK/) PARKER W B.
PA (WAUD/) WAUD W.
PA (GADI/) GADI V K.
```

PA	(BENN/)	BENNETT L L.
XX	Sorscher EJ,	Parker WB, Waud W, Gadi VK, Bennett LL;
XX	WPI;	2003-897265/82.
XX	Use of purine analogs for inhibiting tumor growth and treating virally infected cells.	
XX	Example 25;	Page 20-23; 40pp; English.
XX	The present invention relates to a method of inhibiting tumour growth by administration of a purine analogue. The method comprises the stable transformation of a prokaryotic host cell with a vector including a DNA sequence for purine nucleoside phosphorylase (PNP) or hydrolase. The CC transformed prokaryotic host cell is used in combination with a purine CC substrate to treat tumour and/or virally infected cells. The method CC provides toxin gene therapy to overcome the problem of inefficient CC delivery, cell replication dependent killing and low toxin diffusion CC between cells. The present sequence represents the Escherichia coli CC purine nucleoside phosphorylase (PNP) gene.	
XX	Sequence 5013 BP;	1231 A; 1258 C; 1325 G; 1199 T; 0 U; 0 Other;
Qy	Query Match	99.6%; Score 716.8; DB 10; Length 5013;
Db	Best Local Similarity	99.7%; Pred. No. 2.5e-216;
Qy	Matches 718;	Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Qy	1	ATGCTACCCACACATTAATGCAGAAATGGCGATTTTCGCTGACGTAGTTTCGATGCCA 60
Db	413	ATGGCTACCCACACATTAATGCAGAAATGGCGATTTTCGCTGACGTAGTTTCGATGCCA 472
Qy	61	GGCGACCGCTGCGTGGCGAAGTATATTGCTCAAACTTTCCTTGAAGATGCCCGTGAAGTG 120
Db	473	GGCGACCGCTGCGTGGCGAAGTATATTGCTCAAACTTTCCTTGAAGATGCCCGTGAAGTG 532
Qy	121	AACAACTGTCGCGTATGCTGGGCTTCACCGGTACTTACAAAGCCGCAAAATTTCCGTA 180
Db	533	AACAACTGTCGCGTATGCTGGGCTTCACCGGTACTTACAAAGCCGCAAAATTTCCGTA 592
Qy	181	ATGGGTACCGTGTGGTATCCCGTCTGCTCATCTACACCAAGAACTGATCACCGAT 240
Db	593	ATGGGTACCGTGTGGTATCCCGTCTGCTCATCTACACCAAGAACTGATCACCGAT 652
Qy	241	TTCCGCGTGAAGAAAAATATCCGCGTGGGTTCCTGTGGCGCAGATTCTGCCGCAAGTAAA 300
Db	653	TTCCGCGTGAAGAAAAATATCCGCGTGGGTTCCTGTGGCGCAGATTCTGCCGCAAGTAAA 712
Qy	301	CTGCGCAGCGTCGTTATCGGTATGGGTGCTGTCACCGGATTCCTCAAGTTAACCGCATCCGT 360
Db	713	CTGCGCAGCGTCGTTATCGGTATGGGTGCTGTCACCGGATTCCTCAAGTTAACCGCATCCGT 772
Qy	361	TTTAAAGACCATGACTTTGCGCTATCGCTGACACTTCGACATGGTTCGTAACGCGATGATAG 420
Db	773	TTTAAAGACCATGACTTTGCGCGCTATCGCTGACACTTCGACATGGTTCGTAACGCGATGATAG 832
Qy	421	GCAGCTAAAGCACTGGGTATTGATGCTCGCTGGGTAACTGTTCTTCGGCTGACCTGTTTC 480
Db	833	GCAGCTAAAGCACTGGGTATTGATGCTCGCTGGGTAACTGTTCTTCGGCTGACCTGTTTC 892
Qy	481	TACTCTCCGACCGCGAAAATGTTTCGACGTGATGGAAAAATACGGCATTTCTCGCGGTGAAA 540
Db	893	TACTCTCCGACCGCGAAAATGTTTCGACGTGATGGAAAAATACGGCATTTCTCGCGGTGAAA 952
Qy	541	ATGGAACGGCTGATCTACCGCGTCTGTCAGAAATTTGGCGCGAAGCCCTGACCATC 600
Db	953	ATGGAACGGCTGATCTACCGCGTCTGTCAGAAATTTGGCGCGAAGCCCTGACCATC 1012
Qy	601	TGCACCGTATCTGACCAACATCCGCACTACAGAGCAGACCACTGCGCGCTGACGCTCAGACT 660
Db	1013	TGCACCGTATCTGACCAACATCCGCACTACAGAGCAGACCACTGCGCGCTGACGCTCAGACT 1072
Qy	661	ACCTTCAACGACATGATCAAAATCGCACTGGAATCCGTTCTGCTGGGCGATTAAGAGGTAA 720

||||| 1073 ACCTTCAACGACATGATCAAAATCGCACTGGAATCCGTTCTTCTGCTGGCGGATAAGAGTAA 1133

RESULT 11
 AAAS1633
 ID ID AAAS1633 standard; DNA; 5495 BP.
 XX AC AC AAAS1633;
 XX XX
 DT DT 31-OCT-2000 (first entry)
 XX XX
 DE DE Plasmid pGM707 containing deoD and tetracycline resistance genes.
 XX XX
 KW KW Uridine phosphorylase; udp; purine nucleoside phosphorylase; PNP; deod;
 KW KW Genetically modified organism; catalyst; transglycosylation; anti-viral;
 KW KW nucleoside analogue; alpha-pentose-1-phosphate sugar; phosphorolysis;
 KW KW anti-tumour; mesophilic bacterium; tetracycline resistance; ss.
 XX XX
 OS OS Escherichia coli.
 OS OS Synthetic.
 OS OS Chimeric.
 XX XX
 FH FH Key Location/Qualifiers
 FT FT misc_RNA 1. .230
 FT FT /tag= a
 FT FT /label= pUC18_sequence
 FT FT 216. .952
 FT FT /tag= b
 FT FT /product= "lacZ-udp_fused_gene"
 FT FT 961. .978
 FT FT /tag= c
 FT FT /label= pUC18_sequence
 FT FT 979. .1422
 FT FT /tag= d
 FT FT /label= pHP45_sequence
 FT FT 1423. .2822
 FT FT /tag= e
 FT FT /label= pBR322_Tet_gene_sequence
 FT FT 2823. .3090
 FT FT /tag= f
 FT FT /label= pHP45_sequence
 FT FT 3091. .5495
 FT FT /tag= g
 FT FT /label= pUC18_sequence
 XX XX
 PN PN WO200039307-A2.
 XX XX
 PD PD 06-JUL-2000.
 XX XX
 PP PP 23-DEC-1999; 99WO-EP010416.
 XX XX
 PR PR 23-DEC-1998; 98IT-MI002792.
 XX XX
 PA PA (NORP-) NORPHARMA SPA.
 XX XX
 PI PI Bestetti G, Cali' S, Ghisotti D, Oraini G, Tonon G, Zuffi G;
 XX XX
 DR DR WPI; 2000-452402/39.
 XX XX
 PT PT Recombinant expression vectors used to transform cells for the production
 PT PT of e.g. nucleosides encode uridine phosphorylase and/or purine nucleoside
 PT PT phosphorylase.
 XX XX
 PS PS Claim 14; Page 48-50; 72pp; English.
 XX XX
 CC CC Host cells genetically modified to express uridine phosphorylase (Udp)
 CC CC and purine nucleoside phosphorylase (PNP) or their corresponding crude or
 CC CC purified extracts, either separately or in combination are used as
 CC CC catalysts of transglycosylation reactions between a donor nucleoside and
 CC CC an acceptor base, for preparing nucleoside analogues containing
 CC CC heterocyclic systems with purine and/or pyrimidine bases substituted by
 CC CC one or more nitrogen atoms, for preparing alpha-pentose-1-phosphate

sugars by phosphorolysis reactions and for producing nucleosides and modified nucleoside analogues (all claimed). The modified or natural nucleosides are used directly or as intermediates in the preparation of drugs with anti-viral and anti-tumour activity and for preparing oligonucleotides for therapeutic or diagnostic use. The recombinant host cells are produced using recombinant plasmid expression vectors comprising at least one gene sequence of a mesophilic bacterium coding for a polypeptide with UDP activity and/or enzyme PNP activity and at least one gene sequence coding for antibiotic resistance. In particular, the E. coli udp and deoD genes are used. Recombinant strains produced using the vectors express polypeptides with enzyme UDP and PNP activity in large amounts, e.g. 340-1040 times higher UDP activity and 120-200 times higher PNP activity than non-transformed corresponding wild type strains

XX Sequence 5495 BP; 1335 A; 1416 C; 1460 G; 1284 T; 0 U; 0 Other;

Query Match 99.6%; Score 716.8; DB 3; Length 5495;
Best Local Similarity 99.7%; Pred. No. 2.6e-216;
Matches 718; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 ATGGCTACCCACACATTAATGCGAATGGGCGATTTCGCTGACGTAGTTTGATGCCA 60
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
240 ATGGCTACCCACACATTAATGCGAATGGGCGATTTCGCTGACGTAGTTTGATGCCA 299
Qy 61 GCGGACCGCTGCGTGGCAAGTATATGCTCAAACTTTCCTTGAAGATGCCGGAAGTG 120
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
300 GCGGACCGCTGCGTGGCAAGTATATGCTCAAACTTTCCTTGAAGATGCCGGAAGTG 359
Qy 121 AACACGTTCCGCGTATGCTCGGCTTACCGGTACTTACAAAGCCGCAAAATTTCCGTA 180
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
360 AACACGTTCCGCGTATGCTCGGCTTACCGGTACTTACAAAGCCGCAAAATTTCCGTA 419
Qy 181 ATGGGTACCGTGTGGTATCCGTCCTGCTCCATCTACACCAAGAACGTATCACCAGT 240
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
420 ATGGGTACCGTATGGGTATCCCGTCTGCTCCATCTACACCAAGAACGTATCACCAGT 479
Qy 241 TTCGCGTGAAGAAATTTATCCGCGTGGTTCCTGTGCGCAGTTCGCGCAGTAAAA 300
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
480 TTCGCGTGAAGAAATTTATCCGCGTGGTTCCTGTGCGCAGTTCGCGCAGTAAAA 539
Qy 301 CTGCGCGACGTCGTTATCGGTATGGGTGCTGACCGATTCACAAAGTTAACCGCATCCGT 360
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
540 CTGCGCGACGTCGTTATCGGTATGGGTGCTGACCGATTCACAAAGTTAACCGCATCCGT 599
Qy 361 TTTAAAGACCATGACTTTCGCGCTATCGTCACATTCGACATGGTGCCTAAGCAGTAGAT 420
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
600 TTTAAAGACCATGACTTTCGCGCTATCGTCACATTCGACATGGTGCCTAAGCAGTAGAT 659
Qy 421 GCAGTAAAGCACCTGGTATTGATGCTCGCTGGGTAACTGTTCTCGCTGACCTGTTTC 480
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
660 GCAGTAAAGCACCTGGTATTGATGCTCGCTGGGTAACTGTTCTCGCTGACCTGTTTC 719
Qy 481 TACTCTCGGACGGCGAAATGTTCCAGCTGTATGGAATAACGGCATTCGCGCGTGGAA 540
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
720 TACTCTCGGACGGCGAAATGTTCCAGCTGTATGGAATAACGGCATTCGCGCGTGGAA 779
Qy 541 ATGGAAGCGCTGGTATCTACGGGTGCTCGATTCAGAAATTTGCGCGGAAGCCCTGACCATC 600
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
780 ATGGAAGCGCTGGTATCTACGGGTGCTCGATTCAGAAATTTGCGCGGAAGCCCTGACCATC 839
Qy 601 TGCACCGTATCTGACCATCCGACTCAGCAGCAGCAGCAGTCCGCTGAGCGTCACT 660
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
840 TGCACCGTATCTGACCATCCGACTCAGCAGCAGCAGCAGTCCGCTGAGCGTCACT 899
Qy 661 ACCTTCAACGACATGATCAAAATCGCACTGGAATCCGTTCTGCTGGCGGATAAAGAGTAA 720
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
900 ACCTTCAACGACATGATCAAAATCGCACTGGAATCCGTTCTGCTGGCGGATAAAGAGTAA 959

RESULT 12
AAA51635
ID AAA51635 standard; DNA; 6301 BP.

XX AAA51635;
AC
XX 31-OCT-2000 (first entry)
XX
XX Plasmid pGM716 containing udp, deoD and tetracycline resistance genes.
XX
XX Uridine phosphorylase; udp; purine nucleoside phosphorylase; PNP; deoD;
XX Genetically modified organism; catalyst; transglycosylation; anti-viral;
XX nucleoside analogue; alpha-pentose-1-phosphate sugar; phosphorolysis;
XX anti-tumour; mesophilic bacterium; tetracycline resistance; ss.
XX Escherichia coli.
OS Synthetic.
OS Chimeric.
XX
FH Key Location/Qualifiers
FT misc_RNA 243..1021
FT /*tag= a
FT /*label= udp_gene
FT misc_RNA 1037..1766
FT /*tag= b
FT /*label= deoD_gene
FT misc_RNA 2229..3628
FT /*tag= c
FT /*label= pBR322_Tet_gene_sequence
XX
XX W0200039307-A2.
XX
XX 06-JUL-2000.
XX
XX 23-DEC-1999; 99WO-EP010416.
XX
XX 23-DEC-1998; 98IT-MI002792.
XX
XX (NORP-) NORPHARMA SPA.
XX
XX Beatetti G, Cali' S, Ghisotti D, Orsini G, Tonon G, Zuffi G;
XX WPI; 2000-452402/39.
XX
XX Recombinant expression vectors used to transform cells for the production
XX of e.g. nucleosides encode uridine phosphorylase and/or purine nucleoside
XX phosphorylase.
XX
XX Claim 14; Page 53-55; 72pp; English.
XX
XX Host cells genetically modified to express uridine phosphorylase (Udp)
XX and purine nucleoside phosphorylase (PNP) or their corresponding crude or
XX purified extracts, either separately or in combination are used as
XX catalysts of transglycosylation reactions between a donor nucleoside and
XX an acceptor base, for preparing nucleoside analogues containing
XX heterocyclic systems with purine and/or pyrimidine bases substituted by
XX one or more nitrogen atoms, for preparing alpha-pentose-1-phosphate
XX sugars by phosphorolysis reactions and for producing nucleosides and
XX modified nucleoside analogues (all claimed). The modified or natural
XX nucleosides are used directly or as intermediates in the preparation of
XX drugs with anti-viral and anti-tumour activity and for preparing
XX oligonucleotides for therapeutic or diagnostic use. The recombinant host
XX cells are produced using recombinant plasmid expression vectors
XX comprising at least one gene sequence of a mesophilic bacterium coding
XX for a polypeptide with UDP activity and/or enzyme PNP activity and at
XX least one gene sequence coding for antibiotic resistance. In particular,
XX the E. coli udp and deoD genes are used. Recombinant strains produced
XX using the vectors express polypeptides with enzyme UDP and PNP activity
XX in large amounts, e.g. 340-1040 times higher UDP activity and 120-200
XX times higher PNP activity than non-transformed corresponding wild type
XX strains
XX
SQ Sequence 6301 BP; 1507 A; 1625 C; 1686 G; 1483 T; 0 U; 0 Other;

Query Match 99.6%; Score 716.8; DB 3; Length 6301;
Best Local Similarity 99.7%; Pred. No. 2.8e-216;

		Matches	718;	Conservative	0;	Mismatches	2;	Indels	0;	Gaps	0;
QY	1	ATGGCTACCCACACATTAATGAGAAATGGCGGATTTTCGCTGACGTAGTTTGGATGCCA	60								
Db	1046	ATGGCTACCCACACATTAATGAGAAATGGCGGATTTTCGCTGACGTAGTTTGGATGCCA	1105								
QY	61	GGCGACCGCTCGCTGCGAAGTATATGCTGAAACTTTCCTTGAAGATGCCCGTGAAGTG	120								
Db	1106	GGCGACCGCTCGCTGCGAAGTATATGCTGAAACTTTCCTTGAAGATGCCCGTGAAGTG	1165								
QY	121	AACAAAGCTTCCGGGTATGCTGGGCTTACCGGTACTTACAAAGCGCGAAATTTCCGTA	180								
Db	1166	AACAAAGCTTCCGGGTATGCTGGGCTTACCGGTACTTACAAAGCGCGAAATTTCCGTA	1225								
QY	181	ATGGGTACGGGTGTTGGTATCCCGTCTGCTCCATCTACACAAAGAACTGATCACCGAT	240								
Db	1236	ATGGGTACGGGTATGGGTATCCCGTCTGCTCCATCTACACAAAGAACTGATCACCGAT	1285								
QY	241	TTGGCGGTGAAGAAATTAATCCCGTGGGTTCTGTGGGCGAGTTCTGCGGCACGTAAAA	300								
Db	1286	TTGGCGGTGAAGAAATTAATCCCGTGGGTTCTGTGGGCGAGTTCTGCGGCACGTAAAA	1345								
QY	301	CTGGCGGACGCTCGTTATCGGTATGGGTGCTGCGACCGATTTCCAAAGTTAAACCGATCCGT	360								
Db	1346	CTGGCGGACGCTCGTTATCGGTATGGGTGCTGCGACCGATTTCCAAAGTTAAACCGATCCGT	1405								
QY	361	TTTAAAGACCATGACTTTTGGCGCTATCGCTGACTTCGACATGGTGGTAAACGCACTAGAT	420								
Db	1406	TTTAAAGACCATGACTTTTGGCGCTATCGCTGACTTCGACATGGTGGTAAACGCACTAGAT	1465								
QY	421	GCAGCTAAGCACTGGGTATTGATGCTCGCGTGGTAAACCTGTTCTCGCTGACCTGTTTC	480								
Db	1466	GCAGCTAAGCACTGGGTATTGATGCTCGCGTGGTAAACCTGTTCTCGCTGACCTGTTTC	1525								
QY	481	TACTCTCCGGAGCGGAAATGTTCCAGCTGATGGAAATACCGCATTTCTGGCGTGGAA	540								
Db	1526	TACTCTCCGGAGCGGAAATGTTCCAGCTGATGGAAATACCGCATTTCTGGCGTGGAA	1585								
QY	541	ATGGAAGCGGCTGGTATCTACGGCTCGCTGAGAAATTTGGCGGAAAGCCCTGACCATC	600								
Db	1586	ATGGAAGCGGCTGGTATCTACGGCTCGCTGAGAAATTTGGCGGAAAGCCCTGACCATC	1645								
QY	601	TGCACCGTATCTGACACATCCGCACTCAGAGCAGACACATGCGCTGAGCGTCAGACT	660								
Db	1646	TGCACCGTATCTGACACATCCGCACTCAGAGCAGACACATGCGCTGAGCGTCAGACT	1705								
QY	661	ACCTTCAACGACATCATCAAAATCCACTGGAATCCGTTCTGCTGGGCGATAAAGATAA	720								
Db	1706	ACCTTCAACGACATCATCAAAATCCACTGGAATCCGTTCTGCTGGGCGATAAAGATAA	1765								
RESULT 13											
AAD56238											
ID	AAD56238 standard; DNA; 720 BP.										
XX											
AC	AAD56238;										
XX											
DT	07-AUG-2003 (first entry)										
XX											
DE	Escherichia coli mutant PNP enzyme (A157V) encoding DNA #SEQ ID 3.										
XX											
KW	Purine nucleoside phosphorylase; mutant; tumour; PNP; enzyme; cancer;										
KW	leukaemia; glioblastoma; lymphoma; melanoma; carcinoma; gene therapy;										
KW	gene; de.										
XX											
OS	Escherichia coli.										
OS	Synthetic.										
FH	Key										
XX											
FT	Location/Qualifiers										
XX											
FT	1..720										
XX											
FT	/*tag= a										
XX											
FT	/product= "Mutant PNP enzyme"										
XX											

PN	W02003035012-A2.
XX	
PD	01-MAY-2003.
XX	
PF	28-OCT-2002; 2002WO-US034626.
XX	
PR	26-OCT-2001; 2001US-00035300.
XX	
PA	(UABR-) UAB RES FOUND.
PA	(SOUR-) SOUTHERN RES INST.
XX	(CORR-) CORNELL RES FOUND INC.
PI	Elalick SE, Parker WB, Seer1st JA, Sorscher BJ;
XX	
DR	WPI; 2003-421350/39.
DR	N-PSDB; AAE37218.
XX	
PT	Novel purified mutant purine cleaving enzyme having different biological activity than a wild-type purine cleaving enzyme, useful for treating a pathological condition characterized by abnormal cell growth.
PT	
XX	Claim 12; Page 65-66; 67pp; English.
PS	
XX	The invention relates to mutant purine nucleoside phosphorylase (PNP) enzymes and nucleic acid molecules encoding such enzymes. These mutant enzymes have greater biological activity than wild-type enzymes. PNP enzymes are useful for treating a pathological condition characterised by abnormal cell growth such as that occurring in cancer of the skin.
CC	breast, bone, colon, head or neck, kidney, larynx, lung, liver, ovary, nasopharynx, oesophagus, pancreas, prostate, bladder, rectum, stomach, thyroid, testicle and uterus as well as other conditions characterised by abnormal cell growth such as myeloid leukaemia, glioblastoma and lymphoma. Sequences of the invention are useful for treating metastatic solid tumours such as melanoma, pancreatic, liver or colonic carcinoma or to activate prodrugs in the treatment of cancer. They are useful for treating virally infected cells. They are also useful in gene therapy.
CC	The present sequence is Escherichia coli mutant PNP enzyme (A157V) encoding DNA
XX	
SQ	Sequence 720 BP; 172 A; 189 C; 187 G; 172 T; 0 U; 0 Other;

Query Match		99.3%;	Score 715.2;	DB 8;	Length 720;
Best Local Similarity		99.6%;	Pred. No. 3.1e-216;		
Matches		717;	Conservative	0;	Mismatches 3;
				Indels	0;
				Gaps	0;
QY	1	ATGGCTACCCACACATTAATGAGAAATGGCGGATTTTCGCTGACGTAGTTTGGATGCCA	60		
Db	1	ATGGCTACCCACACATTAATGAGAAATGGCGGATTTTCGCTGACGTAGTTTGGATGCCA	60		
QY	61	GGCGACCGCTCGCTGCGAAGTATATGCTGAAACTTTTCTTGAAGATGCCCGTGAAGTG	120		
Db	61	GGCGACCGCTCGCTGCGAAGTATATGCTGAAACTTTTCTTGAAGATGCCCGTGAAGTG	120		
QY	121	AACAAAGCTTCCGGGTATGCTGGGCTTACCGGTACTTACAAAGCGCGAAATTTCCGTA	180		
Db	121	AACAAAGCTTCCGGGTATGCTGGGCTTACCGGTACTTACAAAGCGCGAAATTTCCGTA	180		
QY	181	ATGGGTACGGGTATGGGTATCCCGTCTGCTCCATCTACACAAAGAACTGATCACCGAT	240		
Db	181	ATGGGTACGGGTATGGGTATCCCGTCTGCTCCATCTACACAAAGAACTGATCACCGAT	240		
QY	241	TTGGCGGTGAAGAAATTAATCCCGTGGGTTCTGTGGCGAGTTCTGCCGACGTAAAA	300		
Db	241	TTGGCGGTGAAGAAATTAATCCCGTGGGTTCTGTGGCGAGTTCTGCCGACGTAAAA	300		
QY	301	CTGGCGGACGCTGTTATCGGTATGGGTGCTGACCGGATTTCAAAGTTAAACCGATCCGT	360		
Db	301	CTGGCGGACGCTGTTATCGGTATGGGTGCTGACCGGATTTCAAAGTTAAACCGATCCGT	360		
QY	361	TTTAAAGACCATGACTTTTGGCGCTATCGCTGACTTCGACATGGTGGTAAACGCACTAGAT	420		
Db	361	TTTAAAGACCATGACTTTTGGCGCTATCGCTGACTTCGACATGGTGGTAAACGCACTAGAT	420		

QY 421 GCAGCTAAAGCAGCTGGGTATTGATGCTCGCGTGGGTAAACCTGTTCTCCGCTGACCTGTTTC 480
DB |||||||
QY 421 GCAGCTAAAGCAGCTGGGTATTGATGCTCGCGTGGGTAAACCTGTTCTCCGCTGACCTGTTTC 480
DB |||||||
QY 481 TACTCTCCGGACGGCGAAATGTTTCGACGTGATGGAATAACGGCATTCCTCGGCGTGGAA 540
DB |||||||
QY 541 ATGGAAGCGGTGATATCTACGGGTGCTGCGAATTTGGCGGAAGCCCTGACCATC 600
DB |||||||
QY 601 TGCACCGTATCTGACCACTCCGCACTCACGAGCAGACCACTGCGCGTGCAGCTCAGACT 660
DB |||||||
QY 661 ACCTTCAACGACATGATCAAAATCGCACTGGAATCCGTTCTGTGGCGGATAAAGAGTAA 720
DB |||||||

RESULT 14

AAF55446
ID AAF55446 standard; DNA; 720 BP.

AC AAF55446;

DT 29-MAY-2001 (first entry)

DE Nucleotide sequence of a purine nucleoside phosphorylase.

XX Deoxyribonucleoside; deoxyribose 1-phosphate; thymidine phosphorylase;
KW purine nucleoside phosphorylase; phosphopentose mutase;
KW phosphopentose aldolase; fructose 1,6-diphosphate aldolase;
KW deoxyribokinase; nucleoside 2-deoxyribose-1-phosphate aldolase; ss.
XX Escherichia coli.

XX Key Location/Qualifiers

FT CDS 1..720

FT /*tag= a

FT /product= "purine nucleoside phosphorylase"

XX WO200114566-A2.

PD 01-MAR-2001.

XX 18-AUG-2000; 2000WO-EP008088.

XX 20-AUG-1999; 99EP-00116425.

XX (HOFF) ROCHE DIAGNOSTICS GMBH.

XX (INSP) INST PASTEUR.

XX (PHAR-) PHARMA-WALDHOFF GMBH & CO KG.

XX Tischer W, Ihlenfeldt H, Barzu O, Sakamoto H, Pistotnik E;

PI Marliere P, Pochet S;

XX WPI; 2001-235026/24.

DR P-PSDB; AAB67590.

XX In vitro enzymatic synthesis of deoxyribonucleosides comprises reacting
PT deoxyribose 1-phosphate and a nucleobase to form a deoxyribonucleoside
PT and an inorganic phosphate.

XX Disclosure; Page 63-64; 73pp; English.

XX The present sequence encodes a purine nucleoside phosphorylase enzyme.
CC This enzyme is involved in the biosynthesis of deoxyribonucleosides, and
CC is used in the method of the invention. The specification describes a
CC method for the in vitro enzymatic synthesis of deoxyribonucleosides. The
CC method comprises reacting deoxyribose 1-phosphate and a nucleobase to
CC form a deoxyribonucleoside and an inorganic phosphate. Enzymes which may

CC be used in the method of the invention include thymidine phosphorylase,
CC purine nucleoside phosphorylase, phosphopentose mutase, phosphopentose
CC aldolase, fructose 1,6-diphosphate aldolase, deoxyribokinase, and
CC nucleoside 2-deoxyribose-1-phosphate aldolase

XX SQ Sequence 720 BP; 174 A; 190 C; 185 G; 171 T; 0 U; 0 Other;

Query Match 99.1%; Score 713.6; DB 5; Length 720;

Best Local Similarity 99.4%; Pred. No. 9.9e-216;

Matches 716; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 ATGGCTACCCACACATTAAATGAGAAATGGGCGATTTTCGCTGACGTAGTTTGTATGCCA 60

DB 1 ATGGCTACCCACACATTAAATGAGAAATGGGCGATTTTCGCTGACGTAGTTTGTATGCCA 60

QY 61 GGGACCCCGTGGTGGAGATATATGCTGAAACTTTCCTTGAAGTCCCGTGAAGTG 120

DB 61 GGGACCCCGTGGTGGAGATATATGCTGAAACTTTCCTTGAAGTCCCGTGAAGTG 120

QY 121 AACAAAGTTTCGCGGTATGCTGGGCTTTCACCGGTACTTACAAAGCGCGCAAAATTTCCGTA 180

DB 121 AACAAAGTTTCGCGGTATGCTGGGCTTTCACCGGTACTTACAAAGCGCGCAAAATTTCCGTA 180

QY 181 ATGGGTACAGGTCTGGTATCCCGTCTGCTCCATCTACACAAAGAACTGATCAGCAT 240

DB 181 ATGGGTACAGGTCTGGTATCCCGTCTGCTCCATCTACACAAAGAACTGATCAGCAT 240

QY 241 TTGCGGCTGAGAAATATATCCCGTGGGTTCCTGTGGCGCAGTTCTGCGCGACGTAAAA 300

DB 241 TTGCGGCTGAGAAATATATCCCGTGGGTTCCTGTGGCGCAGTTCTGCGCGACGTAAAA 300

QY 301 CTGCGGACGCTCGTATCGGTATGGGTGCTGCACCGGATTCCTCAAGTTAACCGCATCCGT 360

DB 301 CTGCGGACGCTCGTATCGGTATGGGTGCTGCACCGGATTCCTCAAGTTAACCGCATCCGT 360

QY 361 TTTAAAGACCATGACTTTTGGCGCTATCGCTGATCTCGAATGTGTGCGTAAACGAGTAGAT 420

DB 361 TTTAAAGACCATGACTTTTGGCGCTATCGCTGATCTCGAATGTGTGCGTAAACGAGTAGAT 420

QY 421 GCAGCTAAAGCACTGGGTATTGATGCTCGCGTGGGTAAACCTGTTCTCCGCTGACCTGTTTC 480

DB 421 GCAGCTAAAGCACTGGGTATTGATGCTCGCGTGGGTAAACCTGTTCTCCGCTGACCTGTTTC 480

QY 481 TACTCTCCGGACGGCGAAATGTTTCGACGTGATGGAATAATACCGCATTTCTCGCGCTGGAA 540

DB 481 TACTCTCCGGACGGCGAAATGTTTCGACGTGATGGAATAATACCGCATTTCTCGCGCTGGAA 540

QY 541 ATGGAAGCGGCTGGTATCTACGGCGTGGTGCAGAAATTTTGGCGGAAAGCCCTGACCATC 600

DB 541 ATGGAAGCGGCTGGTATCTACGGCGTGGTGCAGAAATTTTGGCGGAAAGCCCTGACCATC 600

QY 601 TGCACCGTATCTGACCACTCCGCACTCACGAGCAGACCACTGCGCGTGAAGTGAAGTAA 660

DB 601 TGCACCGTATCTGACCACTCCGCACTCACGAGCAGACCACTGCGCGTGAAGTGAAGTAA 660

QY 661 ACCTTCAACGACATGATCAAAATCGCACTGGAATCCGTTCTGTGGCGGATAAAGAGTAA 720

DB 661 ACCTTCAACGACATGATCAAAATCGCACTGGAATCCGTTCTGTGGCGGATAAAGAGTAA 720

RESULT 15

AAA51641

ID AAA51641 standard; DNA; 3066 BP.

XX AAA51641;

AC AAA51641;

XX 31-OCT-2000 (first entry)

XX pGM747 containing deob without upstream promoter.

XX Uridine phosphorylase; udp; purine nucleoside phosphorylase; PNP; deob;
KW Genetically modified organism; catalyst; transglycosylation; anti-viral;
KW nucleoside analogue; alpha-pentose-1-phosphate sugar; phosphorolysis;

anti-tumour; mesophilic bacterium; tetracycline resistance; ss.

Escherichia coli.

Synthetic.

Chimeric.

WO200039307-A2.

06-JUL-2000.

23-DEC-1999; 99WO-EP010416.

23-DEC-1998; 98IT-MI002792.

(NORP-) NORPHARMA SPA.

Beetetti G, Cali' S, Ghisotti D, Orsini G, Tonon G, Zuffi G;

WPI; 2000-452402/39.

Recombinant expression vectors used to transform cells for the production of e.g. nucleosides encode uridine phosphorylase and/or purine nucleoside phosphorylase.

Claim 14; Page 66-67; 72pp; English.

Host cells genetically modified to express uridine phosphorylase (Udp) and purine nucleoside phosphorylase (PNP) or their corresponding crude or purified extracts, either separately or in combination are used as catalysts of transglycosylation reactions between a donor nucleoside and an acceptor base, for preparing nucleoside analogues containing heterocyclic systems with purine and/or pyrimidine bases substituted by one or more nitrogen atoms, for preparing alpha-pentose-1-phosphates sugars by phosphorolysis reactions and for producing nucleosides and modified nucleoside analogues (all claimed). The modified or natural nucleosides are used directly or as intermediates in the preparation of drugs with anti-viral and anti-tumour activity and for preparing oligonucleotides for therapeutic or diagnostic use. The recombinant host cells are produced using recombinant plasmid expression vectors comprising at least one gene sequence of a mesophilic bacterium coding for a polypeptide with Udp activity and/or enzyme PNP activity and at least one gene sequence coding for antibiotic resistance. In particular, the E. coli udp and deob genes are used. Recombinant strains produced using the vectors express polypeptides with enzyme Udp and PNP activity in large amounts, e.g. 340-1040 times higher Udp activity and 120-200 times higher PNP activity than non-transformed corresponding wild type strains

Sequence 3066 BP; 746 A; 767 C; 772 G; 746 T; 0 U; 35 Other;

Query Match 88.9%; Score 639.8; DB 3; Length 3066;

Best Local Similarity 98.8%; Pred. No. 5.6e-192;

Matches 718; Conservative 0; Mismatches 2; Indels 7; Gaps 7;

Qy	1	ATGCTACCCACACATTAATGACAGAAATGGCGATTTCGCTGACGTAGTTTTCATGCCA	60
Db	102	ATGCTACCCACACATTAATGACAGAAATGGCGATTTCGCTGACGTAGTTTTCATGCCA	161
Qy	61	GGCGACCGCTCGCTGCGAAGTATATTCCTGAAATTTCC-TTGAAGTCCCGTGAAGT	119
Db	162	GGCGACCGCTCGCTGCGAAGTATATTCCTGAAATTTCCNTTGAAGTCCCGTGAAGT	221
Qy	120	GAACACCGTTCCGCTATGCTGGGCTTCACCGGTACTTACAAAGCCGCAAAATTTCCGT	179
Db	222	GAACACCGTTCCGCTATGCTGGGCTTCACCGGTACTTACAAAGCCGCAAAATTTCCGT	281
Qy	180	AATGGGTACCGTGTGGTAT-CCCGTCTGCTCCATCTACACCAAGAACTGATCACCG	238
Db	282	AATGGGTACCGTATGGGTATCCCGTCTGCTCCATCTACACCAAGAACTGATCACCG	341
Qy	239	ATTTCCGGGTGACAGAAATTAATCCCGTGGGTTCTGCTGGCGCAGTTTCGCCGACGTAA	298
Db	342	ATTTCCGGGTGACAGAAATTAATCCCGTGGGTTCTGCTGGCGCAGTTTCGCCGACGTAA	401

Qy	299	AA-CTGCGCGACGTCGTTATCGGTATCGGTGCTGCACCGATTCCAAAGTTAACCGCATC	357
Db	402	AANCTGCGCGACGTCGTTATCGGTATCGGTGCTGCACCGATTCCAAAGTTAACCGCATC	461
Qy	358	CGTTTTAAAGACCATGACTTTGCGCTATCGCTGACTTCGACA-TGGTGGCTTAACGCAGT	416
Db	462	CGTTTTAAAGACCATGACTTTGCGCTATCGCTGACTTCGACA-TGGTGGCTTAACGCAGT	521
Qy	417	AGATGCAGCTAAAGCACTGGGTATTGATGCTCGGTGGGTAAACCTGTTCTCCGCTGACCT	476
Db	522	AGATGCAGCTAAAGCACTGGGTATTGATGCTCGGTGGGTAAACCTGTTCTCCGCTGACCT	581
Qy	477	GTTCTACTCTCCGGACCGGGAAT-GTTTCGACGTGATGGAATAATACGGCATTTCTCGCG	535
Db	582	GTTCTACTCTCCGGACCGGGAATGTTTCGACGTGATGGAATAATACGGCATTTCTCGCG	641
Qy	536	TGGAATGGAAGCGGCTGCTATCTACGGCTCGCTGCAGAAATTTGGCGCGAAAGCCCTGA	595
Db	642	TGGAATGGAAGCGGCTGCTATCTACGGCTCGCTGCAGAAATTTGGCGCGAAAGCCCTGA	701
Qy	596	CCATC-TGCACCGGTATCTGACCACTCCGCACTCAGCAGCAGACCACTGCCGCTGAGCGT	654
Db	702	CCATC-TGCACCGGTATCTGACCACTCCGCACTCAGCAGCAGACCACTGCCGCTGAGCGT	761
Qy	655	CAGACTACTTCAACGACATGATCAAAATCGCACTGGAATCCGTTTC-TGCTGGGCGATAA	713
Db	762	CAGACTACTTCAACGACATGATCAAAATCGCACTGGAATCCGTTTC-TGCTGGGCGATAA	821
Qy	714	AGAGTAA 720	
Db	822	AGAGTAA 828	

Search completed: July 22, 2005, 19:37:30
Job time : 727 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 22, 2005, 17:50:12 ; Search time 229 Seconds
(without alignments)
5144.625 Million cell updates/sec

Title: US-10-035-300A-1
Perfect score: 720
Sequence: 1 atgctaccacacattaa.....tgctggcgataaagagtaa 720

Scoring table: IDENTITY NUC
Gapop 10,0 , Gapext 1.0

Searched: 1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents NA.*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PTUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	716.8	99.6	5013	US-09-183-188B-5	Sequence 5, Appli
2	558.4	77.6	732	US-09-489-039A-1631	Sequence 1631, Ap
3	420.8	58.4	795	US-09-543-681A-2371	Sequence 2371, Ap
C 4	402.6	55.9	1830121	US-09-557-884-1	Sequence 1, Appli
C 5	402.6	55.9	1830121	US-09-643-990A-1	Sequence 1, Appli
6	265	36.8	744	US-09-902-540-3568	Sequence 3568, Ap
C 7	265	36.8	19019	US-09-902-540-1171	Sequence 1171, Ap
C 8	243.4	33.8	640681	US-09-790-988-1	Sequence 1, Appli
9	210.8	29.3	5301	US-08-956-171E-443	Sequence 443, App
10	210.8	29.3	5301	US-08-781-986A-443	Sequence 443, App
11	204.2	28.4	711	US-09-583-110-37	Sequence 37, Appl
12	204.2	28.4	714	US-09-107-433-1967	Sequence 1967, Ap
13	204.2	28.4	13440	US-08-961-527-128	Sequence 128, App
14	198.6	27.6	735	US-09-107-532A-1185	Sequence 1185, Ap
15	180.8	25.1	2962	US-09-710-279-3511	Sequence 3511, Ap
16	179.2	24.9	732	US-09-134-001C-1402	Sequence 1402, Ap
17	174.6	24.2	714	US-09-134-000C-2608	Sequence 2608, Ap
18	157.4	21.9	2006	US-08-956-171E-378	Sequence 378, App
19	157.4	21.9	2006	US-08-781-986A-378	Sequence 378, App
C 20	141.4	19.6	5739	US-09-634-238-1	Sequence 1, Appli
21	103.4	14.8	1124	US-08-809-254A-3	Sequence 3, Appli
22	99.4	13.8	580073	US-08-545-528D-1	Sequence 1, Appli
23	39.8	5.5	483	US-09-252-991A-7028	Sequence 7028, Ap
C 24	39.8	5.5	2016	US-09-252-991A-7070	Sequence 7070, Ap
25	39.8	5.5	6855	US-09-252-991A-6976	Sequence 6976, Ap
26	37.6	5.2	1830121	US-09-557-884-1	Sequence 1, Appli
27	37.6	5.2	1830121	US-09-643-990A-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1

US-09-183-188B-5
; Sequence 5, Application US/09183188B
; Patent No. 6491905
; GENERAL INFORMATION:
; APPLICANT: Sorscher, Eric J.
; APPLICANT: Parker, William B.
; APPLICANT: Waud, William
; APPLICANT: Gadi, Vijaykrishna K.
; TITLE OF INVENTION: RECOMBINANT E. COLI FOR DELIVERY OF PNP TO TUMOR CELLS
; FILE REFERENCE: UAB-12405/22
; CURRENT APPLICATION NUMBER: US/09/183,188B
; CURRENT FILING DATE: 1998-10-30
; PRIOR APPLICATION NUMBER: 08/702,181
; PRIOR FILING DATE: 1996-08-23
; PRIOR APPLICATION NUMBER: 08/122,321
; PRIOR FILING DATE: 1993-09-14
; PRIOR APPLICATION NUMBER: 60/064,676
; PRIOR FILING DATE: 1997-10-31
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 5013
; LENGTH: 5013
; TYPE: DNA
; ORGANISM: Escherichia coli
US-09-183-188B-5

Query Match 99.6%; Score 716.8; DB 4; Length 5013;
Best Local Similarity 99.7%; Pred. No. 1.3e-225;
Matches 718; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy	1	ATGCTACCCACACATTAATGCGAAGTATTCCTGCTGAGATGCCGCGTGAAGTG	60
Db	413	ATGCTACCCACACATTAATGCGAAGTATTCCTGCTGAGATGCCGCGTGAAGTG	472
Qy	61	GCGACCGCGTGGCGAAGTATTCCTGCTGAGATGCCGCGTGAAGTG	120
Db	473	GCGACCGCGTGGCGAAGTATTCCTGCTGAGATGCCGCGTGAAGTG	532
Qy	121	AACACGTTCCGGTATGCTGGGCTTACCGGTACTTACAAAGCGCGAAAATTTCCGTA	180
Db	533	AACACGTTCCGGTATGCTGGGCTTACCGGTACTTACAAAGCGCGAAAATTTCCGTA	592
Qy	181	ATGGGTACCGTGTGGTATCCCGTCTGCTCATCTACCAAGAACTGATACCCGAT	240
Db	593	ATGGGTACCGTGTGGTATCCCGTCTGCTCATCTACCAAGAACTGATACCCGAT	652
Qy	241	TTCCGCGTGAAGAAATTCCTGCGTGGTTCCTGCGCGAGTTCTCCCGCAGCTAAA	300
Db	653	TTCCGCGTGAAGAAATTCCTGCGTGGTTCCTGCGCGAGTTCTCCCGCAGCTAAA	712

Qy	301	CTGCGCAGCGTCGTTATCGGTATGGGTGCTGCAACGATTCCAAAGTTAAACCGCATCCGT	360
Db	713	CTGCGCAGCGTCGTTATCGGTATGGGTGCTGCAACGATTCCAAAGTTAAACCGCATCCGT	772
Qy	361	TTTAAAGACCATGACTTTTGGCGCTATCGCTGACTTCGACATGTTGCGTAACGCAGTAGAT	420
Db	773	TTTAAAGACCATGACTTTTGGCGCTATCGCTGACTTCGACATGTTGCGTAACGCAGTAGAT	832
Qy	421	GCAGCTAAAGCAGCTGGGTATTGATGCTCGGTGGGTAACTGTTCTCCGCTGACCTGTC	480
Db	833	GCAGCTAAAGCAGCTGGGTATTGATGCTCGGTGGGTAACTGTTCTCCGCTGACCTGTC	892
Qy	481	TACTCTCCGAGCGCGAAATGTTTCGACGTGATGGAAAAATACGGCATTTCTCGGCGTGGAA	540
Db	893	TACTCTCCGAGCGCGAAATGTTTCGACGTGATGGAAAAATACGGCATTTCTCGGCGTGGAA	952
Qy	541	ATGGAAACGGCTGTTATCTACGGGCTCGCTGCAGAAATTTGGCGCGAAAGCCCTGACCATC	600
Db	953	ATGGAAACGGCTGTTATCTACGGGCTCGCTGCAGAAATTTGGCGCGAAAGCCCTGACCATC	1012
Qy	601	TGCACCGGTATCTGACCAACATCCGCACATCAGAGCAGACCACTGCGCGCTGAGCGGTCAACT	660
Db	1013	TGCACCGGTATCTGACCAACATCCGCACATCAGAGCAGACCACTGCGCGCTGAGCGGTCAACT	1072
Qy	661	ACCTTCAACGACATGATCAAAATCGCACTCGGAATCCGTTCTGTCTGGCGCGATAAGAGTAA	720
Db	1073	ACCTTCAACGACATGATCAAAATCGCACTCGGAATCCGTTCTGTCTGGCGCGATAAGAGTAA	1132

```

RESULT 2
US-09-489-039A-1631
; Sequence 1631, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489.039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 1631
; LENGTH: 732
; TYPE: DNA
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-1631

```

Query Match	77.6%;	Score 558.4;	DB 4;	Length 732;
Best Local Similarity	86.0%;	Pos. No. 8.3e-174;		
Matches 619;	Conservative 0;	Mismatches 101;	Indels 0;	Gaps 0;
Qy	1	ATGGCTACCCACACATTATATGCAGAAATGGCGGATTTCCGTGACGTAGTCTTTGATGCCA	60	
Db	13	ATGSCAACTCTCACATTAAACGCAGAAATGGCGGATTTCCGTGACGTCTGTTTGATCCG	72	
Qy	61	GGCGACCCGCTGGTGCAGACTATATTGCTCAAACTTTCCCTTGAAGATGCCCGTGAAGTG	120	
Db	73	GGCGACCCGCTGGCGCGAAGACATTTGCAGAAACCTTTCCCTCGAAGACGTGCGTGAAGTG	132	
Qy	121	AACAACTTTCGCGGTATGCTGGGGCTTCACCGGTACTTACAAAGCCGCCAAAATTTCCGTA	180	
Db	133	AACAACTGGCGGCATGCTTGGGGCTTCACCGGTACCTATTAAGGCCCGAAAATCTCCGTG	192	
Qy	181	ATGGGTCAACGGTGTGTGTATCCCGTCTCGTCCATCTACACCAAGAACTGATCACCGAT	240	
Db	193	ATGGGCCAACGGATGGGGATCCCGTCTGCTCTATCTACACCAAGAGCTGATCACCGAT	252	
Qy	241	TTGGCGGTGAAGAAAATTTATCCGGCTGGGTTCTGTGCGCGAGTTCTGCCGCGACGTAAAA	300	
Db	253	TTGGCGGTGAAGAAAATTCATCCGGTGGGGTCTTGGCGCGCGGTGCGTGAAGATGTCAA	312	

Qy	301	CTGCGCAGCGTCGTTATTCGGTATACGGTGCCTGACACCGATTCAAAGTTTAAACCGCATTCGGT	360
Db	313	CTGCGTGACCGTGGTGATCGGCATCGGGCGCGTCGACCGGACTCTAAAGTGAAACCGTCTGCGT	372
Qy	361	TTTTAAAGACCATGACTTTGCGCGCTATCGCTGACACTTCGACATGGTGCCTAAACGACGTAGAT	420
Db	373	TTCAAGACCAATGATTTTGGCGCGATATGCCGATTTTCGGCATGGTGCGGTAATGCGGTAGAC	432
Qy	421	GCAGCTAAAGCACATGGGTATTGATGCTCCGCTGGGTAACTCTTCCTCCGCTGACCTGTTTC	480
Db	433	CGCGCGAAAGCGCTGGCGGTGACCGCGCGCTCGGCACACTCTTCTCCGCAGACCTGTTTC	492
Qy	481	TACTCTCCGAGCGCGGAAATGTTTCGACGTGATGGAATAATACGGCATTTCTCGGCGTGGAA	540
Db	493	TATACGCGCGACCCGCTCCATCTTTCGACGTGATGGAATAATACGGCATTTCTGGCGCTGGAA	552
Qy	541	ATGGAAGCGCTGCTATCTACGGCGTCTGCTGCAGAAATTTGGCGCGAAAGCCCTGACCATC	600
Db	553	ATGGAAGCGCGCGGTATCTACGGCGTGGCGCGGAGTTTCGGCGCGAAAGCGCTGACCATC	612
Qy	601	TGCACCGGTATCTGACCATCCGACACTCACGACGAGACCACTTCGCGCTGAGCGTCCAGACT	660
Db	613	TGCACCGCTCTCCGACCATCCGTTACCCACGACGAGACCACTTCGCGCTGAAACGTCAGACG	672
Qy	661	ACCTTCAACGACATGATCAAAATGCACTGGAATCCGTTCTGCTGGCGCGATAAAGATTA	720
Db	673	ACCTTCAACGACATGATCAAAATCGCGTGGAAATCCGTCGCTGGCGCGATAAAGATTA	732
RESULT 3			
US-09-543-681A-2371			
; Sequence 2371, Application US/09543681A			
; Patent No. 6605709			
; GENERAL INFORMATION:			
; APPLICANT: GARY BRETON			
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEIN			
; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS			
; FILE REFERENCE: 2709.1002-001			
; CURRENT APPLICATION NUMBER: US/09/543,681A			
; PRIOR FILING DATE: 2000-04-05			
; PRIOR APPLICATION NUMBER: US 60/128,706			
; PRIOR FILING DATE: 1999-04-09			
; NUMBER OF SEQ ID NOS: 8344			
; SEQ ID NO 2371			
; LENGTH: 795			
; TYPE: DNA			
; ORGANISM: Proteus mirabilis			
US-09-543-681A-2371			

Query Match	58.4%;	Score 420.8;	DB 4;	Length 795;
Best Local Similarity	74.9%;	Pred. No. 2.3e-128;		
Matches 527;	Conservative 0;	Mismatches 177;	Indels 0;	Gaps 0
Qy	1	ATGCGTACCCACACATTAAATGCGAGAAATGGCGGATTTGCGTGACGTAGTTTTTGATGCCA	60	
Db	79	ATGCGTACCCCTCATATTAAACGCGAGAAATGGCGGATTTGCTGATGTCGTTTAAATGCGG	138	
Qy	61	GGCGACCGCTGCTGCGGAAGTATATGCTCAAAACCTTTCTTGAAGATGCCCGTGAAGTG	120	
Db	139	GGCGACCGCTGCTGTGTAAATACATTGCTGAAACCTTTTTTACAAGACGTACGTCAGTA	198	
Qy	121	AACAAAGTTTCGCGGTATGCTGGGGCTTCACCGGTACTTACAAGGCGCGAAAATTTCCGTA	180	
Db	199	AATTAATGTTGTGCGCATGTTAGTGTTTACTGTGACGTATAAAGGCGGTAAATTTCTGTA	258	
Qy	181	ATGGGTACCGGTGTTGGTATACCCGTCCTGCTCCATCTACACAAAGAACTGATCACCAGT	240	
Db	259	ATGGCCACCGGTA TGGGTATTCCTTCCTGTGCTCAATTTATGCTAAAGAGCTAATTACTGAT	318	
Qy	241	TTCCGCGTGAAGAAAAATTATCCGCGTGGGTTCTGTGCGCGAGTTCCTCCGCGACGTAAAA	300	
Db	319	TTTGGTGTAAAGTGATTTATTCGGGTGGGTTTCATGTGGTGGGTATTACAGATGTTGAA	378	

QY 301 CTGCGCGAGCTGGTATCGGTATGGTGCCTGACCGGATTCGAAAGTTAAACGCGATCGT 360
| | | | |
Db 379 TTACGTGATGTTGATCGGTATGGGAGCATGTACAGATTCTTAAAGTTAAACCGCTTACGC 438
| | | | |
QY 361 TTAAAGACCATGACTTTCGCGCTATCGCTGACTTCGAGATGGTGGTAAACGAGTAGAT 420
| | | | |
Db 439 TTTAAAGACCAAGATTTTTCGCGCTATTCGTGATTTTGAATTAGTACAAAAATGCTGTATCT 498
| | | | |
QY 421 GCAGCTAAAGCACTGGGTATTGATGCTCGCGTGGGTAAACCTGTTCTCCGCTGACCTGTT 480
| | | | |
Db 499 GCTGCGAAAGCAAGATATTAAAGTTGCTGGGTATATTTCTCAGCGGATCTTTTC 558
| | | | |
QY 481 TACTCTCCGAGCGCGAAATGTTTCAGCTGATGGAATAATACGGCAATTCCTGGCGTGGAA 540
| | | | |
Db 559 TACTCGCTGATCCAGAAATGTTGATGTGATGGAAATAACGGCATCTTAGGTGTTGAA 618
| | | | |
QY 541 ATGGAAGCGCTGATCTACGGCGCTGCTGCAGAAATTTGGCGGAAAGCCCTGACCATC 600
| | | | |
Db 619 ATGGAAGCAGCAGGTATTATGTTGTTAGCTGCGGAATATGGTGCACTGCGCTGACAAAT 678
| | | | |
QY 601 TGCACCGTATCGACCATCGCACTCACGAGCAGACACTGCCGCTGAGCGTCAGACT 660
| | | | |
Db 679 TGTACTGTTCTGATCACTTAAGAAAGGTACACAAACACTGAGAAAGCGGCCAACT 738
| | | | |
QY 661 ACCTTCAACGACATGATCAAAATCGCACTGGAATCCGTTCTGCT 704
| | | | |
Db 739 ACCTTTAATGAATGATTGAAATTCGATTAGAACTCTGTTCTGTT 782
| | | | |

RESULT 4

US-09-557-884-1/c
; Sequence 1, Application US/09557884
; Patent No. 6506581
; GENERAL INFORMATION:
; APPLICANT: Fleischmann et al.
; TITLE OF INVENTION: The Nucleotide sequence of
; the Haemophilus influenzae Rd Genome, Fragments
; Thereof, and Uses Thereof

NUMBER OF SEQUENCES: 1

CORRESPONDENCE ADDRESS:

ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: MD
COUNTRY: USA

ZIP: 20850

COMPUTER READABLE FORM:

MEDIUM TYPE: 3 1/2 inch diskette
COMPUTER: Dell Pentium
OPERATING SYSTEM: MS DOS v6.22
SOFTWARE: ASCII Text

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/557,884
FILING DATE: 25-Apr-2000
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/476,102
FILING DATE: JUN-5-1995

ATTORNEY/AGENT INFORMATION:

NAME: Michelle S. Marks
REGISTRATION NUMBER: 41,971
REFERENCE/DOCKET NUMBER: PB186P3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 301-309-8504
TELEFAX: 301-309-8439

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:
LENGTH: 1830121 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 1:

US-09-557-884-1

Query Match 55.9%; Score 402.6; DB 4; Length 1830121;
Best Local Similarity 72.7%; Pred. No. 2.4e-120;
Matches 519; Conservative 0; Mismatches 195; Indels 0; Gaps 0;
QY 7 ACCCCACACATTAATGCGAGAAATGGCGCATTTTCGCTGACGTAGTTTTGATGCCAGCGCAC 66
| | | | |
Db 540876 ACTCCACATATAACCGCGCTGAAAGGTGCATTTGCTGATGTTGTTTAAATGCTGCGCAT 540817
| | | | |
QY 67 CCGCTGCTGCGAAGTATATTGCTGAAACTTTTCCCTTGAAGATGCCCGTGAAGTGAACAAC 126
| | | | |
Db 540816 CCACCTCGTGCAGAAATATATTGCGAGAAACTTTTTTACAGGATGTTGTTGAAGTAAACGAT 540757
| | | | |
QY 127 GTTCGCGTATGCTGGGCTTCACCGGTACTTACAAAGGCCCGCAAAATTTCCGTAATGGGT 186
| | | | |
Db 540756 GTTCGTAATATGCTTGGTTTTTACTGGAACCTTATAAGGTCGTAAAATTTCTATTATGGG 540697
| | | | |
QY 187 CACGGTGTGGTATCCCGTCTCTGCTCATCTACACCAAGAACTGATCACCAGTTTCGGC 246
| | | | |
Db 540696 CACGGTATGGGGAATTCATCTTGTCTATTATTCGCGAAAGAAATTAATCACTGAATATGCG 540637
| | | | |
QY 247 GTGAAGAAATATCCCGTGGGTTCCTGTGCGCAGTTCTGCCGACGTAACAACTGGC 306
| | | | |
Db 540636 GTGAAGAAATATTCGCTGTAGGTTCTTGTGGCACTGTTGCTATGGACGTGAAAGTGGC 540577
| | | | |
QY 307 GACGTGTTATCGGTATGGGTGCTGCACCGGATTCCAAAGTTAAACCGCATCGTTTTAAA 366
| | | | |
Db 540576 GATGTGATTATCGGTTTAGGTGTCATGTACTGATTCAAAAGTAAACCGTATTCTGTTTCAA 540517
| | | | |
QY 367 GACCATGATTTGGCGCTATCGCTGATCTTCGACATGTTGCGTAAACGAGTAGATCGAGCT 426
| | | | |
Db 540516 GATAACGACTTTNCAGCTATTGCTGACTTGTATATGGCACAAAGCGCTGTTCAAGCTGCA 540457
| | | | |
QY 427 AAGCACTGGGTATTGATGCTCGCGTGGGTAACTGTTCTCGCTGACCTGTTCTACTCT 486
| | | | |
Db 540456 AAGCAAAAGTAAAGTCTGTTGTTAGTAAATTTTCTCGCGGATTTATTTCTATACA 540397
| | | | |
QY 487 CCGGACGCGGAAATGTTTCGACGTGATGGAATAATACGGCATTTCTCGCGTGGAAATGGAA 546
| | | | |
Db 540396 CCCGATGTGGAATGTTTCGATGTAATGGAATAATACGGCATTTTAGGTGTGGAATGGAA 540337
| | | | |
QY 547 CCGGCTGATATCTAGCGGCTCGCTGCGAAGTTTGGCGGAAAGCCCTGACCATCTGCACC 606
| | | | |
Db 540336 GCTGCTGGAATTTATGTTGCTGCGCTGCAGAAATATGTTGCAAAAGCATTAACGATTTGTACC 540277
| | | | |
QY 607 GTATCTGACCATCCGCACTCACGAGCAGACCACTGCGCTGAGCGTCAGACTACCTTC 666
| | | | |
Db 540276 GTGCTGACCATATTCGTACTCACGAAACAAACCGGAGAAACGTCATTAACATTC 540217
| | | | |
QY 667 AACGACATGATCAAAATCGCACTGGAATCCGTTCTGCTGGGCGATAAAGAGTAA 720
| | | | |
Db 540216 AATGATATGATTGAAATTCGTTAGATTAGTAAATTTGTTGATGCATATAA 540163
| | | | |

RESULT 5

US-09-643-990A-1/c
; Sequence 1, Application US/09643990A
; Patent No. 6528289

GENERAL INFORMATION:

APPLICANT: Robert D. Fleischmann
Mark D. Adams
Owen White
Hamilton O. Smith
J. Craig Venter

TITLE OF INVENTION: The Nucleotide sequence of
the Haemophilus influenzae Rd Genome, Fragments
Thereof, and Uses Thereof

NUMBER OF SEQUENCES: 1

CORRESPONDENCE ADDRESS:

ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville,

```
; STATE: MD
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3 1/2 inch diskette
; COMPUTER: Dell Pentium
; OPERATING SYSTEM: MS DOS v6.22
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/643,990A
; FILING DATE: 23-Aug-2000
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/487,429
; FILING DATE: 1995-06-07
; APPLICATION NUMBER: 08/426,787
; FILING DATE: 1995-04-21
; ATTORNEY/AGENT INFORMATION:
; NAME: Kenley K. Hoover
; REGISTRATION NUMBER: 40,302
; REFERENCE/DOCKET NUMBER: P8186P1C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 301-610-5790
; TELEFAX: 310-309-8439
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1830121 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-643-990A-1

Query Match 55.9%; Score 402.6; DB 4; Length 1830121;
Best Local Similarity 72.7%; Pred. No. 2.4e-120;
Matches 519; Conservative 0; Mismatches 195; Indels 0; Gaps 0;

QY 7 ACCCCACACATTAATGCAGAAATGGGCGATTCGCTGACGTAGTTTGGATGCCAGGGCAC 66
Db ACTCCACATATTAAACGGCCCTGAAGGTGCATTTGCTGATGTGTTTAAATGCCCTGGCGAT 540817
QY 67 CCGCTGCGTGCAGTAGTATTGCTGAAACTTTTCCTTGAAGATGCCGTGAAGTGAACAAC 126
Db CCATCTGTCGAAATATATTGCAGAAACTTTTTCACAGATGTGTTGTAAGTAAACGAAT 540757
QY 127 GTTCGGCGTATGCTGGCGTTACCGGTACTTACAAAGCGCGCAAAATTCGCTAAATGGGT 186
Db GTTCGTAATATGCTTGGTTTACTGGAACCTTATAAAGGTGCTAAATTTCTATTATGGGG 540697
QY 187 CACGGTGTGGTATCCGCTCTGCTCCATCTACACCAAGAACTGATCACCGATTTTCGGC 246
Db CACGTAATGGGATTCATCTTGCTCTATTACCGGAAGAAATTAATCACTGAATATGGC 540637
QY 247 GTGAAGAAAATTAATCCGCGTGGGTTCCTGTGGCGAGTTCTGCGGCACTGTAAGTGGCG 306
Db GTGAAGAAAATTAATCCGCGTGGGTTCCTGTGGCGAGTTCTGTATGGACGTGTAAGTGGCG 540577
QY 307 GAGTCGTTATCGGTATGGGTGCGCTGACCGATTCGAAAGTTAACCGCATCGTTTAA 366
Db GATGTGATATCGGTTTATGGTGCATGTACTGATTCGAAAGTAAACCGTATTCGTTTCAA 540517
QY 367 GACCATGACATTTGCGGTATCGCTGACTTCGACATGTGTGTAACCGAGTAGATGACGT 426
Db GATAACGACATTTNACGATTTGCTGACTTTGATATGCAACGCGCTGTTCAAGCTGCA 540457
QY 427 AAAGCACTGGGTATGTATGCTCGCGTGGGTAAACCTGTTCTCCGTGACCTGTTTCTACT 486
Db AAAGCAAAAGTAAAGTCGTTGCTGATAGTAAATTTATTTCTCGCGGATTTTATCTATA 540397
QY 487 CCGGACCGCGAAATGTTTCGACGTGATCGGAAAAATACGGCATTCCTCGCGGTGGAAATG 546
Db CCCGATGTGAAATGTTTCGATGTAATGGAATAATACGGCATTTTATGTTGGAAATGGAA 540337
```

```
547 GCGGCTGGTATCTACGCGCTCGCTGCAGAAATTTGGCGGAAAGCCCTGACCATCTGCACC 606
Db GCTGCTGGAAATTTATGGTGGCTGCAGATATATGGTGCAGAAAGCATTAACGATTTGTACC 540277
QY 607 GTATCTGACCAATCCGCACTCAAGAGCAGACCACTGCCCTGAGCGTCAAGCTACCTTC 666
Db GTGCTGACCATATTCGTACTCAGCAACAAACCAACGCGCAGAGAAAGCTCAATTAACATTC 540217
QY 667 AACGACATGATCAAAATCGCACTCGGAATCGTTCTGCTGGCGGATAAAGAGTAA 720
Db AATGATATGATTGAAATTCGTTAGATTTCAGTATTAAATTTGGTATGCACTATTAA 540163

RESULT 6
US-09-902-540-3568
; Sequence 3568, Application US/09902540
; Patent No. 6833447
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15849)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; CURRENT FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/217,883
; PRIOR FILING DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 3568
; LENGTH: 744
; TYPE: DNA
; ORGANISM: Myxococcus xanthus
US-09-902-540-3568

Query Match 36.8%; Score 265; DB 4; Length 744;
Best Local Similarity 61.3%; Pred. No. 6e-77;
Matches 427; Conservative 0; Mismatches 270; Indels 0; Gaps 0;

QY 1 ATGGCTACCCACACATTAATGCAGAAATGGGCGATTCGCTGACGTAGTTTGGATGCCA 60
Db 1 ATGGCGACTCTCATATCTCCGCTGCCCTGGTGACTTCGCTGACGTGGTCTCATGSCCT 60
QY 61 GGGGACCGCTGGTGCAGAGTATATTGCTGAAACTTTTCCTGAAAGATGCCCGTGAAGTG 120
Db 61 GGGGACCGCTGGTGCAGAGTATATTGCTGAAACTTTTCCTGAAAGATGCCCGTGAAGTG 120
QY 121 AACACAGTTTCGCGTATGCTGGGCTTTCACCGGTACTTACAAAGCGCGCAAAATTTCCGTA 180
Db 121 ACCTCGTGGCAACATGCTCGGCTTTCACCGGACCTTCCGGGGCGCGCGTGTCCGTG 180
QY 181 ATGGGTACGCGTGGTGGTATCCGCTCTGCTCCATCTACACCAAGAACTGATCACCAGT 240
Db 181 ATGGGGCACGGCATGGGCGTTCCTCCATCTCCTACGCTTACGCGCCGAGCTCATCAAGAG 240
QY 241 TTGCGGTGAAGAAAATTAATCCGCGTGGGTTCCTGTGGCGCAGTTCTGCGCGCAGTAAAA 300
Db 241 TATGGGTGGCGGTGATCATCCCGTGGGAGCTGCGGCGCTGAGCACCGAGCTGAAG 300
QY 301 CTGCGGACGCTGTTATCGGTATGGGTGCTGCAACCGATTCGAAAGTTAACCGCATCCGT 360
Db 301 GTCCGGAGGTATCATGCTCGGACGGGGCGGCAACGAGCTCCAAACGTTGAATCGATCGG 360
QY 361 TTTAAAGACATGACTTTGCGCTATCGCTGACTTCGACATGTGTGCGTACGACGTAGAT 420
Db 361 CTGATGGGATGACTTCGCGCGGTGGCGAGCTTCAACGCTCGCGCGCGGCGCATGGAG 420
QY 421 GCAGCTAAAGCACTGGGTATTTGATGCTCGGTCGGTAAACCTGTTCTCCGCTGACCTGTT 480
Db 421 GCGCGGAGCGGCGCAACAGCCTGTGCGGCGCGGCGCGCTTTCACCTCCGACCTCTTC 480
QY 481 TACTCTCCGACCGCGCAAAATGTTTCGACGTGATGGAATAATAACGCAATTTCTCGCGGTG 540
```

Db 481 TACCAACCGCAGGAGAGCTCAACGCCACCTCGCGCGGATGGGTGTCTCGCCGTCGAG 540
Qy 541 ATGGAAGCGGCTGGTATCTACGCGCTCGCTGAGAAATTTGGCGGAAAGCCCTGACCAATC 600
Db 541 ATGGAAGCGGCTGGTATCTACGCGCTCGCTGAGAAATTTGGCGGAAAGCCCTGACCAATC 600
Qy 601 TGCACCGTATCTGACCAATCTGACCAATCTGACCAATCTGACCAATCTGACCAATCTGACCAATC 660
Db 601 CTCACGGTGTGGACCAATCTGACCAATCTGACCAATCTGACCAATCTGACCAATCTGACCAATC 660
Qy 661 ACCTTCAACGATCATCAAAATCGCACTGGAATCCG 697
Db 661 ACGTTGACGAGATGATTGAATCGCCCTGACGTCG 697

RESULT 7

US-09-902-540-1171/c
; Sequence 1171, Application US/09902540
; Patent No. 6833447
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15849)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; CURRENT FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/217,883
; PRIOR FILING DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 1171
; LENGTH: 19019
; TYPE: DNA
; ORGANISM: Myxococcus xanthus
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(19019)
; OTHER INFORMATION: unsure at all n locations
US-09-902-540-1171

Query Match 36.8%; Score 265; DB 4; Length 19019;
Best Local Similarity 61.3%; Pred. No. 4.2e-76;
Matches 427; Conservative 0; Mismatches 270; Indels 0; Gaps 0;
Qy 1 ATGGCTACCCACACATTAATGCGAATGCGGATTTTCGCTGAGTAGTTTGTGATGCA 60
Db 16556 ATGGGACTCTCATATCTCGCTGCGCTGGTACTTGGCTGAGTGTCTCATGCT 16497
Qy 61 GCGACCCGCTCGGTGCGAAGTATATGCTGAACTTTCTTGAAGATGCCCGTGAAGTG 120
Db 16496 GCGACCCGCTCGGTGCGAAGTATATGCTGAACTTTCTTGAAGATGCCCGTGAAGTG 16437
Qy 121 AACACGTTCCGGTATCTCGGCTTACCGGTACTTACAAGGCGCAAAATTTCCGTA 180
Db 16436 ACCTCCGGTCAACATGCTCGGCTTACCGGACCTTCCGGGCGCGGCTGTCGCTG 16377
Qy 181 ATGGTCAAGTGTGGTATCCGCTCTGCTCCATCTACCAAGAACTGATCACCAGT 240
Db 16376 ATGGGCAAGTGTGGTATCCGCTCTGCTCCATCTTCACTACGCAAGGCTCATCAGAG 16317
Qy 241 TTCGGCGTGAAGAAATATCCGCTGGGTTCTTGTGGCGAGTTCTGCGCAGCTAA 300
Db 16316 TATGGGTCGCGTATCATCCGCTGGGAGCTGCGGCGCTGAGCACCGAGTGAAG 16257
Qy 301 CTGGCGAGCTGTTATCGGTATGGGTGCTGACCGGATTCGAAATTAACCGATCCGT 360
Db 16256 GTCCGGAGGTCATGCTCGACGCGGCGCGGACGAGTCCAAACGTAATCGGATCGG 16197
Qy 361 TTTAAAGACCATGACTTTTGGCGCTATCGCTGACTTCGACATGGTGGTAAAGCAGTAGAT 420
Db 16196 CTGATGGGCAATGACTTTCCCGCGGTGCGGACTTTCAGCTCGCGCGCGCGCATGGAG 16137

Qy 421 GCAGCTAAGCACTGGGTATTTGATGCTCGGTGGGTAACTGTTTCTCGCTGACCTGTC 480
Db 16136 GCGCGGAGCGCGCAACAGGCTGTGCGCGCGCGCGCTTTCACCTCCGACCTTTC 16077
Qy 481 TACTCTCGGACGCGGAATGTTGACGCTGATGGAATAATACGGCATTTCTCGCGCTGGA 540
Db 16076 TACCACCGCAGGAGAGCTCAACGCCACCTCGCGCGGATGGGTGTCTGCGCCCTCGAG 16017
Qy 541 ATGGAAGCGGCTGGTATCTACGCGCTCGCTGAGAAATTTGGCGGAAAGCCCTGACCAATC 600
Db 16016 ATGGAAGTGGCGGCTCTACGCGCTGCGCGGAGTGGCGCGCGCTGGGGCTG 15957
Qy 601 TGCACCGTATCTGACCAATCTGACCAATCTGACCAATCTGACCAATCTGACCAATCTGACCAATC 660
Db 15956 CTCACGGTGTGGACCAATCTGACCAATCTGACCAATCTGACCAATCTGACCAATCTGACCAATC 15897
Qy 661 ACCTTCAACGATCATCAAAATCGCACTGGAATCCG 697
Db 15896 ACGTTGACGAGATGATTGAATCGCCCTGACGTCG 15860

RESULT 8

US-09-790-988-1/c
; Sequence 1, Application US/09790988
; Patent No. 6632935
; GENERAL INFORMATION:
; APPLICANT: SHIGENOBU, SHUJI
; APPLICANT: WATANABE, HIDEMI
; APPLICANT: HATTORI, MASAHIRA
; APPLICANT: SAKAKI, YOSHIYUKI
; TITLE OF INVENTION: GENOME DNA OF BACTERIAL SYMBIONT OF APHIDS
; FILE REFERENCE: 081356/0159
; CURRENT APPLICATION NUMBER: US/09/790,988
; CURRENT FILING DATE: 2001-02-23
; PRIOR APPLICATION NUMBER: JP2000-107160
; PRIOR FILING DATE: 2000-04-07
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 640681
; TYPE: DNA
; ORGANISM: Buchnera sp.
US-09-790-988-1

Query Match 33.8%; Score 243.4; DB 4; Length 640681;
Best Local Similarity 59.2%; Pred. No. 4.6e-68;
Matches 415; Conservative 0; Mismatches 286; Indels 0; Gaps 0;
Qy 1 ATGGCTACCCACACATTAATGCGAAGTATTTGCTGAAAGTCCCTTGAAGATGCCCGTGAAGTG 120
Db 575671 ATGTCTACTCCACATATTAATAGTAAATAAGATGATTTTTCAGATATAGTTTAAATGCG 575612
Qy 61 GCGACCCGCTCGGTGCGAAGTATTTGCTGAAACTTTCTTGAAGATGCCCGTGAAGTG 120
Db 575611 GGAGACCCGCTCGGTGCGAAGTATTTGCTGAAACTTTCTTGAAGATGCCCGTGAAGTG 575552
Qy 121 AACACGTTTCGCGTATGCTGGGCTTACCCGCTACTTACAAAGCGCGCAAAATTTCCGTA 180
Db 575551 AATGATCTCGTTAAATGTTAGCTTATACCGGATTTTATAAATAAGATTTTCAATC 575492
Qy 181 ATGGTCAAGTGTGGTATCCCGTCTGCTCCATCTACCAAGAACTGATCACCAGT 240
Db 575491 ATGAGTCATGTTAGGAATACCATCAGCTTCTCTATACAGAGATTAATTTGAA 575432
Qy 241 TTCGGCTGAGAAATTTATCCGCTGGGTTCTTGTGGCGAGTTCTGCGCGCAGTAA 300
Db 575431 TTTAATGTAAAAAATTTATTCGTATAGGAATTTTGTGTGCTGTACGAGATGATATAAG 575372
Qy 301 CTGGCGCAGCTGTTATCGGTATCGGTGCTGACCGGATTCGAAATTAACCGATCCGT 360
Db 575371 TTACGTGATATAGTAATCAGTATGGGAGCTTCTACTGATTTCTAAGTAATAGATAGA 575312
Qy 361 TTTAAAGACCATGACTTTTGGCGCTATCGCTGACTTCGACATGGTGGTGGTAAACGAGTAGAT 420

SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/781.986A
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Benson, Bob
REGISTRATION NUMBER: 30,446
REFERENCE/DOCKET NUMBER: PB248PP
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 443:
SEQUENCE CHARACTERISTICS:
LENGTH: 5301 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-08-781-986A-443

Query Match 29.3%; Score 210.8; DB 4; Length 5301;
Best Local Similarity 56.5%; Pred. No. 1.5e-58;
Matches 392; Conservative 0; Mismatches 302; Indels 0; Gaps 0;
Qy 4 GCTACCCACACATTAATGCAGAAATGGCGCATTCGCTGACGTAGTTTTTCATGCCAGGC 63
Db 4097 GGTACACACATATTCACCAAAATGGAGTAATAATGCTAAACAGATTAATATGCTGGC 4156
Qy 64 GACCCGCTGCTGCGAAGTATATTCGTGAATCTTTCCTTGAAGATGCCCGTGAAGTGAAC 123
Db 4157 GATCCGCTACGTCGCAAAATAATTGCTGATAATTTTTAGAAAATGTTGAACAATTTAAC 4216
Qy 124 AACGTTCCGGGTATGCTGGGCTTCACCGGTACTTACAAAGCCGCAAAATTTCCGTAATG 183
Db 4217 GATGATCGTAACATGTTGGTTCACACTGGTACATATAAGGTAAGAAGTTTCTGTAATG 4276
Qy 184 GGTCAACGCTGTGTGATCCCGCTCCTGCTCCATCTACACCAAGAACTGATCACCGATTC 243
Db 4277 GTTCTGGTATGGGTATTCCAAGTATTTGGTATTTACTCATATGATGATATACACTCTTT 4336
Qy 244 GCGGTGAAGAAATATTCGCGTGGGTTCTGTGCGCAGTTCTGCCGACGTAAAACTG 303
Db 4337 GATGTAGATACAAATCATTCGTATCGGTTCTGTGGCGCATTTACAAGAAATGTTAACTTA 4396
Qy 304 CGCAGCTGTTATCGGTATCGGTGCGCTGCACCGCATTCCAAGTTAACCGCATCGTTTT 363
Db 4397 TACGATGTTATTTGCAACAAGCTGCATCAACTAAATTCAAATTTATGTAGATCAATACAT 4456
Qy 364 AAAGACCATGATTTTGGCGCTATCGCTGACITTCGACATTCGATGGTGGTAAAGCAGTAGATGCA 423
Db 4457 ATTCAGGTCAATTTCCGCGCTATCGCTGACITTCGATTTAGTAGTAACCTAAAGCTAAATGTC 4516
Qy 424 GCTAAAGCAGTGGGTATTTGATGCTCGCTGGGTAACTGTTTCTCGCTGACCTGTTCTAC 483
Db 4517 GCTGACCAATTCGCTGCTACTACACAGCTAGTAGCTATTTATCTCTGATACATTTTAC 4576
Qy 484 TCTCGGACGCGGAAATGTTGAGCTGATGAGAAATACGCGATTCCTCGCGGTGGAATG 543
Db 4577 AATGCCGATCCAAATTCATATGATGCTTTGGAAAAAATGGGTATTTTAGGTATCGAAATG 4636
Qy 544 GAAGCGCTGCTATCTACGGGCTCGCTGCAGAAATTTGGCGGAAAGCCCTGACCATCTGC 603
Db 4637 GAATCAGCTGTTTATATTTAAATGCGATTCATGCTGGTAAAAAAGCACTTGGTATTTTC 4696
Qy 604 ACCGTATCTGACCATCCGACATCCAGCAGACACCACTCGCTGAGCGTGCAGACTACC 663
Db 4697 ACAGTAAGTATCATATTTTACGTGACGCAAGCTACTACACCTGAAGAAGCTCAAAATTTCA 4756
Qy 664 TTCAACGACATGATCAAAATCGCACTGGAATCCG 697

Db 4757 TTACACAAATGATGGAAATCGCTTTTAGAAATCG 4790
RESULT 11
US-09-583-110-37
Sequence 37, Application US/09583110
Patent No. 6699703
GENERAL INFORMATION:
APPLICANT: Lynn Doucette-Stamm et al.
TITLE OF INVENTION: Nucleic Acid and Amino Acid Sequences Relating to Streptococcus
TITLE OF INVENTION: Pneumoniae for Diagnostics and Therapeutics
FILE REFERENCE: PATH00-07A
CURRENT APPLICATION NUMBER: US/09/583,110
CURRENT FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/107,433
PRIOR FILING DATE: 1998-06-30
PRIOR APPLICATION NUMBER: US 60/085,131
PRIOR FILING DATE: 1998-05-12
PRIOR APPLICATION NUMBER: US 60/051,553
PRIOR FILING DATE: 1997-07-02
NUMBER OF SEQ ID NOS: 5322
SEQ ID NO 37
LENGTH: 711
TYPE: DNA
ORGANISM: Streptococcus pneumoniae
US-09-583-110-37

Query Match 28.4%; Score 204.2; DB 4; Length 711;
Best Local Similarity 57.0%; Pred. No. 6.9e-57;
Matches 393; Conservative 0; Mismatches 293; Indels 3; Gaps 1;
Qy 13 CACATTAATGCAGAAATGGCGATTTTCGCTGACGTAGTTTTTGATGCCAGGACCCGCTG 72
Db 10 CATATTCGCTGCTACAGAGGTGAAATTCGTGATAAAATTTCTTCTCTGGGGATCTCTT 69
Qy 73 CGTGGAGATATATTCGTGAAACTTTCTTTGAAAGATCCCGTGAAGTGAACAAAGTTTCG 132
Db 70 CGTCTAAGTTTATTCGCGGAGAAATTTCTTGTGTATGCTGTTGTTTAAACGAAATGCGT 129
Qy 133 GGTATGCTGGGCTTCACCGGTACTTACAAAGGCGCAAAATTTCCGTAAATGGGTCAACCGT 192
Db 130 AACATGTTGGTACATCTGCTACTTACAGGGTCAACCGTGTATCTGTCTATGCGAACTGG 189
Qy 193 GTTGGTATCCGCTCCTCTCCATCTACACCAAGAACTGATCAACCGATTTTCGCGGTGAAG 252
Db 190 ATGGGAATGCCATCTATTTTCGATTTATGCGCGTGAGTTAATCGTAGACTACGGTGTGAAG 249
Qy 253 AAAATATCCGCTGGGTTCTGTGGCGCAGTTCTGCGGACGTAAACCTGCGGACGCTC 312
Db 250 AAATTTGATTCGTGTGGGAACCTGCAGGTTCTTTGAATGAAGAGGTTTCATGTTTCGTAATTA 309
Qy 313 GTTATCGGTATGGGTGCTGCACCGATTTCCAAAGCTTAAACCGCATCCGTTTAAAGACCAT 372
Db 310 GTTTGGCGGAGCGGCTGCACCACTCAACATCGTTCGTAATGATGATGCGCCACAGTAC 369
Qy 373 GACTTTCCCGCTATCGCTGACTTCCACATCGTGGGTAAACGACGATAGATGACAGCTAAAGCA 432
Db 370 GATTTTCCACAAATTTGCTAGCTTTTGAATTTGCTTCAAAAGCTTACCATATCGCCAAAGAA 429
Qy 433 CTGGGTATGATGCTCGCGTGGGTAACTGTTTCTCGGTGACCTGTTCTACTCTCTCCGAC 492
Db 430 CTGGGTATGACTACTACGTTGGGAACGTTTGTTCATCTGATGCTCTTTTACTCAAAATTA 489
Qy 493 GGCGAAATGTTTCGACGTGATGGAAAAATACGGCATTTCTCGCGGTGGAAATCGAAAGCGCT 552
Db 490 TTTGAAAGAAATATCGAGCTTGGTAAATGGGAGTCAAGGCTGAGGAAATGGAAGACGCA 549
Qy 553 GGTATCTACGCGCTGCTGACAGAAATTTGGCGGAAAGCCCTGACCATCTGACCCGTATCT 612
Db 550 GCTCTTTTACTATCTTGTGCTGCCAATACCATGTTGATGCGCTAGCTATCATGACCATCTCT 609
Qy 613 GACCACATCCGCACTC---ACGACGACCACTCCGCTGAGCGTACAGTACCTTCAAC 669

Db 610 GATAGCTTGCTCAATCCAGACGAGAACACAACTGCAGAGAACGTCAAAATACCTTCACT 669
QY 670 GACATGATCAAAATCGCACTGGATCCGT 698
Db 670 GATATGATGAAGGTGGTTTGGAAACCTT 698

RESULT 12

US-09-107-433-1967
; Sequence 1967, Application US/09107433
; Patent No. 6800744
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID
; SEQUENCES RELATING TO STREPTOCOCCUS PNEUMONIAE FOR DIAGNOSTIC
; THERAPEUTICS
; NUMBER OF SEQUENCES: 5206
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Walham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: <Unknown>
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: <Unknown>
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,433
; FILING DATE: 30-Jun-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/ 085131
; FILING DATE: May 12, 1998
; APPLICATION NUMBER: 60/051553
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Ariniello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-011
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 1967:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 714 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Streptococcus pneumoniae
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (B) LOCATION 1...714
; SEQUENCE DESCRIPTION: SEQ ID NO: 1967:
US-09-107-433-1967

Query Match 28.48; Score 204.2; DB 4; Length 714;
Best Local Similarity 57.08; Pred. No. 6:9e-57;
Matches 393; Conservative 0; Mismatches 293; Indels 3; Gaps 1;

QY 13 CACATTAATGACAGAAATGGCGATTTTCGCTGACGTAGTTTTGATGCCAGCGCACCGCTG 72
Db 13 CATATTGCTGCTCAGCAGGGTGAATTCGTGATAAAATCTTCTCTCGGGATCCTCTT 72
QY 73 CGTGGCAAGTATATGCTGAAATTTCTTGAAGATGCCGTGAAGTGAACAGGTTCCG 132
Db 73 CGTGCTAAGTTTATGCGGAGAAATTCCTTGGTGTGCTGTTGTTTAAACGAAGTGGCT 132

QY 133 GGTATGTCGGGCTTACCGGTACTTACAAAGGCCGCAAAATTTCCGTATATGGTGCAGGT 192
Db 133 AACATGTTTGGTTACACTGCTGTTACAAAGGTCACCGTGTATCTGTCATGGGAACGGG 192
QY 193 GTTGGTATCCCGTCTGCTCCATCTACACAAAGAACTGATACCCGATTTCCGGCGTGAAG 252
Db 193 ATGGGAATGCCATCTATTTCGATTTATGCGCGTGAGTTAATCTAGACTACCGGTGTGAAG 252
QY 253 AAAATTTATCGCGTGGTTCCTGTGGCGCAGTTCTGCCGCACGTAAATCTGCCGACGTC 312
Db 253 AAATTGATTCGTGTGGNACTGCAGGTTCTTTGAATGAAGAGGTTTCATGTCGTGAATTA 312
QY 313 GTTATCGTATGGTGCCTGCACCGATTCCAAAGTTTAAACCGCATCCGTTTAAAGACCAT 372
Db 313 GTTTTGGCGCAGCGGCTGCAACCACTCAAACATCGTTTGTGTAATGACTGGCCACAGTAC 372
QY 373 GACTTTGCCGCTATCGCTGACTTTTCGACATGTCGTAACGCGTGAACGAGTAGATGCGCTAAAGCA 432
Db 373 GATTTTCCCAAAATTCGTAGCTTTGATTTGCTTGATAAAAGCCTTACCATATCGCCAAAGAA 432
QY 433 CTGGTATTGATGCTCGCGTGGTGAACCTGTTCTCCGCTGACCTGTTCTTACTCTCCGGAC 492
Db 433 CTTGGTATGACTACTCACGTTGGGAACGTTTGTGTCATCTGATGTCCTTTACTCAAATTAC 492
QY 493 GCGAAATGTTCCACGCTGATGGAATAATACGGCATTTCTCGCGCTGGAATGGAAGCGGCT 552
Db 493 TTTGAAAGAAATATCAGCTTGGTAAATGGGAGTCAAGGCTGTGGAATGGAAGCAGCA 552
QY 553 GGTATCTAGCGGCTCGCTGCAAGAAATTTGGCGGAAAGCCCTGACCATCTGCAACCGTATCT 612
Db 553 GCTCTTTACTATCTTCTGCCCAATACCATGTTTGATGCGCTAGCTATCATGACCATCTCT 612
QY 613 GACCACATCCGACTC---ACGAGCAGACCACTGCGCGCTGAGCGTCACTACTCTTCAAC 669
Db 613 GATAGCTTGGTCAATCCAGCAGGAACACAACTGCAGAGAAACGTCAAAATACCTTCACT 672
QY 670 GACATGATCAAAATCGCACTGGAATCCGT 698
Db 673 GATATGATGAAGGTGGTTTGGAAACCTT 701

RESULT 13

US-08-961-527-128
; Sequence 128, Application US/08961527
; Patent No. 6420135
; GENERAL INFORMATION:
; APPLICANT: Charles Kunsch
; TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 391
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/961,527
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P1
; TELECOMMUNICATION INFORMATION:

TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 128:
SEQUENCE CHARACTERISTICS:
LENGTH: 13440 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-08-961-527-128

Query Match 28.4%; Score 204.2; DB 3; Length 13440;
Best Local Similarity 57.0%; Pred. No. 4e-56;
Matches 393; Conservative 0; Mismatches 293; Indels 3; Gaps 1;

Qy 13 CACATTAATGAGAAATGGCGATTTTCGCTGACGTAGTTTGTATGCCAGGCGACCCGCTG 72
Db 10453 CATATTGCTGCTAGCAGGGTGAATTCGTGATAAAATTCCTTCCTGGGATCCTCTT 10512

Qy 73 CGTCGGAAGTATATGCTGAAACTTTCTTGAAGATGCCGTGAAGTGAACAAAGTTGCG 132
Db 10513 CGTCTAAGTTTATTCGGGAGAAATTCCTTGATGATGCTGTTGTTTAAACGAAGTCGT 10572

Qy 133 GGTATGCTGGCTTCACCGGTACTTACAAAGCGCAAAATTTCCGTAATGGTCAACGGT 192
Db 10573 AACATGTTGTTACATGCTGACTTACAGGGTCACTGTGATCTGTCATGGGAACCTGG 10632

Qy 193 GTTGGTATCCGCTCCTGCTCATCTACACCAAGAACTGATCAACGATTTTCGGCGTGAAG 252
Db 10633 ATCGGAATGCCATCTATTTCGATTTATCGCGGTGAGTTAATCGTAGACTACGGTGTGAG 10692

Qy 253 AAATTAATCCGCGTGGTTCCTGTGGGCGAGTTCGCGGCAAGTAAACCTCGCGGAGTGC 312
Db 10693 AAATTAATGCTGTGGGAACTGCAGGTTCTTTGAATGAAGAGGTTTCATGTTGTTGAATTA 10752

Qy 313 GTTATCGGTATGGTGTGCTGACCGATTCGAGTTTCAAGTAAACCGATCCGTTTAAAGACAT 372
Db 10753 GTTGTGGCGAGGCGGCTGCAACCACTCAAAACATCGTTTCGTAATGACTGGCCACAGTAC 10812

Qy 373 GACTTTGCCGCTATCGCTGACTTCGACATGTTGTTGTAACGAGTAGATGACGCTTAAAGCA 432
Db 10813 GATTTTCCACAAATTTGCTAGCTTTGATTTGCTTGAATGAGCTACCATATCGCCAAAAA 10872

Qy 433 CTGGGTATGATGCTCGGTGGGTAACTGTTCTTCGCTGACCTGTTCTACTCTCCGGAC 492
Db 10873 CTTGGTATGACTACTCACTGTTGGGAAGCTTTTGTCTGATGCTTTTACTCAAAATAC 10932

Qy 493 GCGAATGTTGACGCTGATGGAATAATACGCAATCTCGCGTGGAAATGGAAGCGGCT 552
Db 10933 TTTGAAAGAAATATCGAGCTTGGTAAATGGGAGTCAAGGCTGTGGAATGGAAGCAGCA 10992

Qy 553 GGTATCTACGGCGTCCGCTGCAGAAATTTGGCGGAAAGCCCTGACCATCTGCACCGTATCT 612
Db 10993 GCTCTTTACTATCTTGTGCCAATACCATGTTGATGGCTAGCTATCATGACCATCTCT 11052

Qy 613 GACCACATCCGCACTC---ACGACGACACACTCGCGTGGAGCTGACACTACCTTCAAC 669
Db 11053 GATAGCTTGGTCAATCCAGACGAGACACAACTGCAGAAGAACTGCAAAATACCTTCACT 11112

Qy 670 GACATGATCAAAATCGCACTGGAATCCGT 698
Db 11113 GATATGATGAGGTTGGTTTGGAAACCTT 11141

RESULT 14
US-09-107-532A-1185
Sequence 1185, Application US/09107532A
Patent No. 6583275
GENERAL INFORMATION:
APPLICANT: Lynn A Doucette-Stamm and David Bush
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
NUMBER OF SEQUENCES: 7310
CORRESPONDENCE ADDRESS:

ADDRESSER: GENOME THERAPEUTICS CORPORATION
STREET: 100 Beaver Street
CITY: Waltham
STATE: Massachusetts
COUNTRY: USA
ZIP: 02354
COMPUTER READABLE FORM:
MEDIUM TYPE: CD-ROM ISO9660
COMPUTER: PC
OPERATING SYSTEM: <Unknown>
SOFTWARE: ASCII
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/107,532A
FILING DATE: 30-Jun-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/085,598
FILING DATE: 14 May 1998
APPLICATION NUMBER: 60/051571
FILING DATE: July 2, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Ariniello, Pamela Deneke
REGISTRATION NUMBER: 40,489
REFERENCE/DOCKET NUMBER: GTC-012
TELECOMMUNICATION INFORMATION:
TELEPHONE: (781)893-5007
TELEFAX: (781)893-8277
INFORMATION FOR SEQ ID NO: 1185:
SEQUENCE CHARACTERISTICS:
LENGTH: 735 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: circular
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Enterococcus faecium
FEATURE:
NAME/KEY: misc feature
LOCATION: (B) LOCATION 1...735
SEQUENCE DESCRIPTION: SEQ ID NO: 1185:
US-09-107-532A-1185

Query Match 27.6%; Score 198.6; DB 4; Length 735;
Best Local Similarity 55.6%; Pred. No. 5e-55;
Matches 381; Conservative 0; Mismatches 304; Indels 0; Gaps 0;

Qy 13 CACATTAATGAGAAATGGCGATTTTCGCTGACGTAGTTTGTATGCCAGGCGACCCGCTG 72
Db 37 CATATTGAAGCGAAAGAGCGGAGATCGCGATTAAGATTTTACTTCCGGGAGATCCTTTG 96

Qy 73 CGTCGGAAGTATATTGCTGAAACTTTCTTGAAGATGCCGTGAAGTGAACAAAGTTTCGC 132
Db 97 CGAGCTAAATATATTGGGAGACTTTTGGAGATCCAGTTTGTACATCAGGTACGC 156

Qy 133 GGTATGCTGGCTTCACCGTACTTACAAAGCGCAAAATTTCCGTAATGGGTACCGT 192
Db 157 GGTATGTTAGCTTTTACCGCAATAATAAAGGAGAAAGTATTTCTGTTCAGGAACAGGA 216

Qy 193 GTTGGTATCCGCTCCTGCTCCATCTACCAAGAACTGATCACCAGATTTTCGGGTGAAG 252
Db 217 ATGGGAATGCCATCAGCCACCATCTATGCCCATGAATTTGATCCAGTCGTATGGTGAAG 276

Qy 253 AAAAATTATCCGCTGGGTTCCTGTGGCGCAGTTCTCGCGCACGTAAACAACTCGCGACGTC 312
Db 277 AAATTGATTCAGTAGGTACATGCGGTGCATTGTCAAAGACGTCCTATTCGTGATTAA 336

Qy 313 GTTATCGGTATGGGTGCTCCACCGATTCACAAAGTTAAACCGCATCCGTTTTAAAGACCAT 372
Db 337 GTACTACCAAGGTGCGAGCGACGAGTTCTTCGATGATCGAAAAAATTTTCAAAGCCTTT 396

Qy 373 GACTTTCGCGTATCGCTGACTTCGACATGGTGGTGAACGAGTAGATGCGAGTAAAGCA 432

Db 397 CATTTCTCTCCAACTACGTGATTTCAATCTATTACTGAAGCATATGAGATAGCTAAAGAA 456
QY 433 CTGGTATTGATCTCGCGTGGGTAACTCTCTCGCTGACCTGTCTACTCTCCGGAC 492
Db 457 AAGGGCTATACGGTTTCATGTAGGAATGTTCTTTCGGAAGATTTCTTCTACAAGATGAT 516
QY 493 GGGGAAATGTTTCGACGTGATGGAATAATACGGCAATCTCTCGCGCTGGAAATGGAAGCGGCT 552
Db 517 TTGACAGAGACGTTCCAAATTAGCTGAACTTGGCGTTTATAGAGTGGAAATGGAGCAGCG 576
QY 553 GGTATCTACGGCGTGTGTCAGAAATTTGGCGGAAAGCCCTGACCACTCTGCACCGCTATCT 612
Db 577 GCTCTTTATTAATAGCGCGGAAATACCATGTGCAGACATTAAGTTTGTGATGACCGTTAGC 636
QY 613 GACCAATCCGCACTCACGAGCAGACACCTGCGCTGAGCGTCAGACTACCTTCAACGAC 672
Db 637 GACCAATTTGATACCCGGTGAAGAAACAAACAGCTGCAGAACGTCAATCTACGTTCAATGAA 696
QY 673 ATGATCAAAATCGCACTGGAATCCG 697
Db 697 ATGATCGAGTAGGATTTGGAAACAG 721

RESULT 15

US-09-710-279-3511
; Sequence 3511, Application US/09710279
; Patent No. 6703492
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/09/710,279
; CURRENT FILING DATE: 2000-11-09
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3511
; LENGTH: 2962
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: nucleic acid sequence
US-09-710-279-3511

Query Match 25.1%; Score 180.8; DB 4; Length 2962;
Best Local Similarity 53.9%; Pred. No. 8.6e-49;
Matches 371; Conservative 0; Mismatches 317; Indels 0; Gaps 0;
QY 6 TACCCACACATTAATGCAGAAATGGCGATTTCGCTGACGTAGTTTGTGATGCCAGCGCA 65
Db 1640 TACACCTCACATTAACCCAAATGGAGTTAAATAGCTTAAACTGTATTAATGCCAGCGCA 1699
QY 66 CCGCGTGTGTCGGAAGTATATTGCTGAAACTTTCTTGAAGATGCCCGTGAAGTGAACAA 125
Db 1700 CCCATTACGTGCTCAATATATTGCTGAAATTTTATAGAAATGTTGAACAATTCATATC 1759
QY 126 CGTTCGGGTATGCTGGGCTTACCGGTACTTACAAAGGCCGCAAAATTTCCGTAATGGG 185
Db 1760 AGTACGTAAATATGTTGGTTTACACAGGAACCTTATAAGGCAAAAGATTTCTGTGATGG 1819
QY 186 TCACGGTGTGTTGATCCCGTCTGCTCCATCTACACCAAGAACTGATCACCGGATTTCCG 245
Db 1820 CTCCTGGAGTGGAAATCCAAAGTATGGTATTTATTTCTATGAACTTACCAATTTCTTGA 1879
QY 246 CGTGAAGAAATTAATCCGCGTGGGTTCTGTGGCGAGTTTCTCGCGCACGTAAAACTGCG 305
Db 1880 TGTGGATACAATCATTCGCGTAGGTTCTGTCGCGCTCTTCAAGAGAGATGTCATTTATA 1939
QY 306 CGAGGTGTTATCGGTATGGGTGCTGCACCGATTCCAAAGTTAACCGGATCCGTTTTAA 365
Db 1940 TGATGTGATTTATTTGCACAAGCCGCTTCCACAAACTCAAAATTTATGTTGATCAATTTAATAT 1999

QY 366 AGACCATGACTTTTGGCGCTATCGCTGACTTCCGACATGGTGCCTAAACGACGTAGATGCAGC 425
Db 2000 ACCAGGTCACTTTTGGCGCTATTGCTGATATATATCTAGTAGCTAAAGCTAAGAAAGTGC 2059
QY 426 TAAAGCACTGGGTATTGATGCTCGCGTGGGTAACTGTCTCTCGCTGACCTGTCTTACTC 485
Db 2060 TGATGAAATTTGGTGCCTATATCATGTAGGTAACTTTTGTCTCATCTGATACATTTTATAA 2119
QY 486 TCCGGACGGGAAATGTTTCGACGTGATGGAATAATACGGCAATTTCTCGGCGTGGAAATGGA 545
Db 2120 TGCTGATTCAACTTTTAAATGATTCATGGAATAAGATGGGTATCTTAGGTATCGAAATGGA 2179
QY 546 AGCGGCTGGTATCTACGGCGTCTGCAGAAATTTGGCGCGAAAGCCCTGACCATCTGCAC 605
Db 2180 ATCTGAGGTCTATATTTAAATGCCATCCATGCAAAATAAAGGCTTTAGGTATTTTAC 2239
QY 606 CGTATCTGACCACTCCGCACTCACGAGCAGACCACTGCGCGTGAAGCGTCAGACTACCTT 665
Db 2240 AGTAAGTGTATCATATTTCTAAGAGATGAAGCCACTAGTGCAGAAAGAGAGACAAACATCAT 2299
QY 666 CAACGACATGATCAAAATCGCACTGGAA 693
Db 2300 TACTCAATGATGGAATAAGCGCTCGAA 2327

Search completed: July 22, 2005, 22:31:51

Job time : 242 secs


```
QY 61 GGGACCCGCTGCGTGCAGAGTATATGCTGAAACTTCTTCTGAAGATGCCCGTGAAGTG 120
DB 61 GGGACCCGCTGCGTGCAGAGTATATGCTGAAACTTCTTCTGAAGATGCCCGTGAAGTG 120
QY 121 AACAAAGTTCCGCGTATGCTGGGCTTCAACCGGTATTTACAAAGCCGCAAAAATTTCCGTA 180
DB 121 AACAAAGTTCCGCGTATGCTGGGCTTCAACCGGTATTTACAAAGCCGCAAAAATTTCCGTA 180
QY 181 ATGGGTACACCGTGTGGTATCCGCTGCTGCTCCATCTACACCAAGAAAGTATGATCACCGAT 240
DB 181 ATGGGTACACCGTGTGGTATCCGCTGCTGCTCCATCTACACCAAGAAAGTATGATCACCGAT 240
QY 241 TTCCGCGTGAAGAAATTTATCCGCGTGGGTTCTGTCGGCGAGTTCTGCGCGCAGTAAAA 300
DB 241 TTCCGCGTGAAGAAATTTATCCGCGTGGGTTCTGTCGGCGAGTTCTGCGCGCAGTAAAA 300
QY 301 CTGCGGACGTCGTTATCGGTATGGGTGCTGCAACCGATTTCCAAAGTTAACCGCATCCGT 360
DB 301 CTGCGGACGTCGTTATCGGTATGGGTGCTGCAACCGATTTCCAAAGTTAACCGCATCCGT 360
QY 361 TTTAAAGACCATGACTTTTGGCGTATCGGTATGGGTGCTGCAACCGATTTCCAAAGTTAACCGCATGAT 420
DB 361 TTTAAAGACCATGACTTTTGGCGTATCGGTATGGGTGCTGCAACCGATTTCCAAAGTTAACCGCATGAT 420
QY 421 GCAGCTAAAGCATCTGGGTATTTGATGCTCGCGTGGGTAACTGTTCTCCGCTGACCTGTTTC 480
DB 421 GCAGCTAAAGCATCTGGGTATTTGATGCTCGCGTGGGTAACTGTTCTCCGCTGACCTGTTTC 480
QY 481 TACTCTCCGACGCGGAAATTTTCGACGTGATGGAATAATACGCGATTTCTCGGCGTGGAA 540
DB 481 TACTCTCCGACGCGGAAATTTTCGACGTGATGGAATAATACGCGATTTCTCGGCGTGGAA 540
QY 541 ATGGAAGCGCTGCTATCTACGCGCTGCTGCAAGATTTTGGCGGAAGCCCTGACCATC 600
DB 541 ATGGAAGCGCTGCTATCTACGCGCTGCTGCAAGATTTTGGCGGAAGCCCTGACCATC 600
QY 601 TGACCCGTATCTGACCATCTCCGACATCAGCAGCAGACCACTGCGCTGAGCGTCAAGCT 660
DB 601 TGACCCGTATCTGACCATCTCCGACATCAGCAGCAGACCACTGCGCTGAGCGTCAAGCT 660
QY 661 ACCTTCAACGACATGATCAAAATCGCATCTGGAATCCGTTCTGCTGGCGGATAAAGAGTAA 720
DB 661 ACCTTCAACGACATGATCAAAATCGCATCTGGAATCCGTTCTGCTGGCGGATAAAGAGTAA 720
```

RESULT 2

```
US-10-891-096-3
; Sequence 3, Application US/10891096
; Publication No. US20050074857A1
; GENERAL INFORMATION:
; APPLICANT: Araki, Tadaashi
; APPLICANT: Miyake, Hitoki
; APPLICANT: Oikawa, Toshihiro
; TITLE OF INVENTION: Method for Producing a Pyrimidine Nucleoside Compound
; TITLE OF INVENTION: and a New Pyrimidine Nucleoside Compound
; FILE REFERENCE: 018765-161
; CURRENT APPLICATION NUMBER: US/10/891,096
; CURRENT FILING DATE: 2004-07-15
; PRIOR APPLICATION NUMBER: JP 2003-199175
; PRIOR FILING DATE: 2003-07-18
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 720
; TYPE: DNA
; ORGANISM: Escherichia coli
US-10-891-096-3
```

Query Match 99.6%; Score 716.8; DB 21; Length 720;
Best Local Similarity 99.7%; Pred. No. 3e-228;
Matches 716; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ATGGCTACCCACACATTAATGCAGAAATGGGCGATTTCGTCGACGTAGTTTTGATGCCA 60

```
DB 1 ATGGCTACCCACACATTAATGCAGAAATGGGCGATTTCGTCGACGTAGTTTTGATGCCA 60
QY 61 GGGACCCGCTGCGTGCAGAGTATATGCTGAAACTTCTTCTGAAGATGCCCGTGAAGTG 120
DB 61 GGGACCCGCTGCGTGCAGAGTATATGCTGAAACTTCTTCTGAAGATGCCCGTGAAGTG 120
QY 121 AACAAAGTTCCGCGTATGCTGGGCTTCAACCGGTATTTACAAAGCCGCAAAAATTTCCGTA 180
DB 121 AACAAAGTTCCGCGTATGCTGGGCTTCAACCGGTATTTACAAAGCCGCAAAAATTTCCGTA 180
QY 181 ATGGGTACACCGTGTGGTATCCGCTGCTGCTCCATCTACACCAAGAAAGTATGATCACCGAT 240
DB 181 ATGGGTACACCGTGTGGTATCCGCTGCTGCTCCATCTACACCAAGAAAGTATGATCACCGAT 240
QY 241 TTCCGCGTGAAGAAATTTATCCGCGTGGGTTCTGTCGGCGAGTTCTGCGCGCAGTAAAA 300
DB 241 TTCCGCGTGAAGAAATTTATCCGCGTGGGTTCTGTCGGCGAGTTCTGCGCGCAGTAAAA 300
QY 301 CTGCGGACGTCGTTATCGGTATGGGTGCTGCAACCGATTTCCAAAGTTAACCGCATCCGT 360
DB 301 CTGCGGACGTCGTTATCGGTATGGGTGCTGCAACCGATTTCCAAAGTTAACCGCATCCGT 360
QY 361 TTTAAAGACCATGACTTTTGGCGTATCGGTATGGGTGCTGCAACCGATTTCCAAAGTTAACCGCATGAT 420
DB 361 TTTAAAGACCATGACTTTTGGCGTATCGGTATGGGTGCTGCAACCGATTTCCAAAGTTAACCGCATGAT 420
QY 421 GCAGCTAAAGCATCTGGGTATTTGATGCTCGCGTGGGTAACTGTTCTCCGCTGACCTGTTTC 480
DB 421 GCAGCTAAAGCATCTGGGTATTTGATGCTCGCGTGGGTAACTGTTCTCCGCTGACCTGTTTC 480
QY 481 TACTCTCCGACGCGGAAATTTTCGACGTGATGGAATAATACGCGATTTCTCGGCGTGGAA 540
DB 481 TACTCTCCGACGCGGAAATTTTCGACGTGATGGAATAATACGCGATTTCTCGGCGTGGAA 540
QY 541 ATGGAAGCGCTGCTATCTACGCGCTGCTGCAAGATTTTGGCGGAAGCCCTGACCATC 600
DB 541 ATGGAAGCGCTGCTATCTACGCGCTGCTGCAAGATTTTGGCGGAAGCCCTGACCATC 600
QY 601 TGACCCGTATCTGACCATCTCCGACATCAGCAGCAGACCACTGCGCTGAGCGTCAAGCT 660
DB 601 TGACCCGTATCTGACCATCTCCGACATCAGCAGCAGACCACTGCGCTGAGCGTCAAGCT 660
QY 661 ACCTTCAACGACATGATCAAAATCGCATCTGGAATCCGTTCTGCTGGCGGATAAAGAGTAA 720
DB 661 ACCTTCAACGACATGATCAAAATCGCATCTGGAATCCGTTCTGCTGGCGGATAAAGAGTAA 720
```

RESULT 3

```
US-09-891-865A-12
; Sequence 12, Application US/09891865A
; Publication No. US20030059870A1
; GENERAL INFORMATION:
; APPLICANT: NORPHARMA SPA
; TITLE OF INVENTION: Recombinant bacterial strains for the production of
; TITLE OF INVENTION: natural nucleosides and modified analogues thereof
; FILE REFERENCE: 99DC26E
; CURRENT APPLICATION NUMBER: US/09/891,865A
; CURRENT FILING DATE: 2002-06-25
; PRIOR APPLICATION NUMBER: MI98A002792
; PRIOR FILING DATE: 1998-12-23
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 12
; LENGTH: 3031
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: udp and deod
; OTHER INFORMATION: cloned into pGM746 without upstream ptac promoter
US-09-891-865A-12
```

Query Match 99.6%; Score 716.8; DB 10; Length 3031;


```

Best Local Similarity 99.7%; Pred. No. 5.7e-228;
Matches 718; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 ATGGCTACCCACACATTAATGCGAATAATGGCGATTTTCGCTGACGTAGTTTTCGATGCCA 60
Db 101 ATGGCTACCCACACATTAATGCGAATAATGGCGATTTTCGCTGACGTAGTTTTCGATGCCA 160

Qy 61 GGCACCCGCTGCTGCGAAGTATATGCTGAAACTTTTCTTTGAAGATGCCCGTGAAGTG 120
Db 161 GGCACCCGCTGCTGCGAAGTATATGCTGAAACTTTTCTTTGAAGATGCCCGTGAAGTG 220

Qy 121 AACAACTTCCGCGTATGCTGGCTTCCCGGTACTTACAAAGCCGCAAAATTTCCGTA 180
Db 221 AACAACTTCCGCGTATGCTGGCTTCCCGGTACTTACAAAGCCGCAAAATTTCCGTA 280

Qy 181 ATGGGTCAACGGTATGCGTATCCCGTCTGCTGCTACACCAAGAACTGATCACCAGT 240
Db 281 ATGGGTCAACGGTATGCGTATCCCGTCTGCTGCTACACCAAGAACTGATCACCAGT 340

Qy 241 TTCCGCTGAAGAAATTAATCCGCTGGGTTCTGTGGCGAGTTCTGCGCACGTAATAA 300
Db 341 TTCCGCTGAAGAAATTAATCCGCTGGGTTCTGTGGCGAGTTCTGCGCACGTAATAA 400

Qy 301 CTGCGCACGCTGTTATCGGTATGCGTATGCGTGGGTAACTGTTCTCGCTGACCTGTC 360
Db 401 CTGCGCACGCTGTTATCGGTATGCGTATGCGTGGGTAACTGTTCTCGCTGACCTGTC 460

Qy 361 TTTAAAGACCATGACTTTGCGCGTATCGTCACTTCGACATGCGTAAACGCGAGTAGAT 420
Db 461 TTTAAAGACCATGACTTTGCGCGTATCGTCACTTCGACATGCGTAAACGCGAGTAGAT 520

Qy 421 GCAGCTAAAGCAGTGGGTATGATGCTGCGTGGGTAACTGTTCTCGCTGACCTGTC 480
Db 521 GCAGCTAAAGCAGTGGGTATGATGCTGCGTGGGTAACTGTTCTCGCTGACCTGTC 580

Qy 481 TACTCTCGGACGCGGAAATGTTTCGACGTATGGAATAATACGGCATTTCTCGGCTGGA 540
Db 581 TACTCTCGGACGCGGAAATGTTTCGACGTATGGAATAATACGGCATTTCTCGGCTGGA 640

Qy 541 ATGGAAGCGCTGCTATCTACGGCGTTCGTCAGAAATTTGCGCGAAGCCCTGACCATC 600
Db 641 ATGGAAGCGCTGCTATCTACGGCGTTCGTCAGAAATTTGCGCGAAGCCCTGACCATC 700

Qy 601 TGCAACCGTATCTGACACATCCGACATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 660
Db 701 TGCAACCGTATCTGACACATCCGACATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 760

Qy 661 ACCTTCAACGACATGATCAAAATCGCACTGGAATCCGTTCTGCTGGGCGATAAAGAGTAA 720
Db 761 ACCTTCAACGACATGATCAAAATCGCACTGGAATCCGTTCTGCTGGGCGATAAAGAGTAA 820

RESULT 4
US-11-036-497-12
; Sequence 12, Application US/11036497
; Publication No. US20050142645A1
; GENERAL INFORMATION:
; APPLICANT: Bestetti, Giuseppina
; APPLICANT: Cali', Simona
; APPLICANT: Ghisotti, Daniela
; APPLICANT: Orsini, Gaetano
; APPLICANT: Tonon, Giancarlo
; APPLICANT: Zuffi, Gabriele
; TITLE OF INVENTION: Recombinant bacterial strains for the production of natural nucle
; TITLE OF INVENTION: and modified analogues thereof
; FILE REFERENCE: 02901/0000410-US0
; CURRENT APPLICATION NUMBER: US/11/036.497
; CURRENT FILING DATE: 2005-01-14
; PRIOR APPLICATION NUMBER: PCT/EP99/10416
; PRIOR FILING DATE: 1999-12-23
; PRIOR APPLICATION NUMBER: MI98A002792
; PRIOR FILING DATE: 1998-12-23
; NUMBER OF SEQ ID NOS: 29

```

```

; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 12
; LENGTH: 3031
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: udp and deoD cloned into pGM746 without upstream ptac promoter
US-11-036-497-12

Query Match 99.6%; Score 716.8; DB 24; Length 3031;
Best Local Similarity 99.7%; Pred. No. 5.7e-228;
Matches 718; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 ATGGCTACCCACACATTAATGCGAATAATGGCGATTTTCGCTGACGTAGTTTTCGATGCCA 60
Db 101 ATGGCTACCCACACATTAATGCGAATAATGGCGATTTTCGCTGACGTAGTTTTCGATGCCA 160

Qy 61 GGCACCCGCTGCTGCGAAGTATATGCTGAAACTTTTCTTTGAAGATGCCCGTGAAGTG 120
Db 161 GGCACCCGCTGCTGCGAAGTATATGCTGAAACTTTTCTTTGAAGATGCCCGTGAAGTG 220

Qy 121 AACAACTTCCGCGTATGCTGGCTTCCCGGTACTTACAAAGCCGCAAAATTTCCGTA 180
Db 221 AACAACTTCCGCGTATGCTGGCTTCCCGGTACTTACAAAGCCGCAAAATTTCCGTA 280

Qy 181 ATGGGTCAACGGTATGCGTATCCCGTCTGCTGCTACACCAAGAACTGATCACCAGT 240
Db 281 ATGGGTCAACGGTATGCGTATCCCGTCTGCTGCTACACCAAGAACTGATCACCAGT 340

Qy 241 TTCCGCTGAAGAAATTAATCCGCTGGGTTCTGTGGCGAGTTCTGCGCACGTAATAA 300
Db 341 TTCCGCTGAAGAAATTAATCCGCTGGGTTCTGTGGCGAGTTCTGCGCACGTAATAA 400

Qy 301 CTGCGCACGCTGTTATCGGTATGCGTATGCGTGGGTAACTGTTCTCGCTGACCTGTC 360
Db 401 CTGCGCACGCTGTTATCGGTATGCGTATGCGTGGGTAACTGTTCTCGCTGACCTGTC 460

Qy 361 TTTAAAGACCATGACTTTGCGCGTATCGTCACTTCGACATGCGTAAACGCGAGTAGAT 420
Db 461 TTTAAAGACCATGACTTTGCGCGTATCGTCACTTCGACATGCGTAAACGCGAGTAGAT 520

Qy 421 GCAGCTAAAGCAGTGGGTATGATGCTGCGTGGGTAACTGTTCTCGCTGACCTGTC 480
Db 521 GCAGCTAAAGCAGTGGGTATGATGCTGCGTGGGTAACTGTTCTCGCTGACCTGTC 580

Qy 481 TACTCTCGGACGCGGAAATGTTTCGACGTATGGAATAATACGGCATTTCTCGGCTGGA 540
Db 581 TACTCTCGGACGCGGAAATGTTTCGACGTATGGAATAATACGGCATTTCTCGGCTGGA 640

Qy 541 ATGGAAGCGCTGCTATCTACGGCGTTCGTCAGAAATTTGCGCGAAGCCCTGACCATC 600
Db 641 ATGGAAGCGCTGCTATCTACGGCGTTCGTCAGAAATTTGCGCGAAGCCCTGACCATC 700

Qy 601 TGCAACCGTATCTGACACATCCGACATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 660
Db 701 TGCAACCGTATCTGACACATCCGACATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 760

Qy 661 ACCTTCAACGACATGATCAAAATCGCACTGGAATCCGTTCTGCTGGGCGATAAAGAGTAA 720
Db 761 ACCTTCAACGACATGATCAAAATCGCACTGGAATCCGTTCTGCTGGGCGATAAAGAGTAA 820

RESULT 5
US-09-891-865A-13
; Sequence 13, Application US/09891865A
; Publication No. US20030059870A1
; GENERAL INFORMATION:
; APPLICANT: NORPHARMA SPA
; TITLE OF INVENTION: Recombinant bacterial strains for the production of
; TITLE OF INVENTION: natural nucleosides and modified analogues thereof
; FILE REFERENCE: 99DC26E
; CURRENT APPLICATION NUMBER: US/09/891.865A
; CURRENT FILING DATE: 2002-06-25

```

```
; PRIOR APPLICATION NUMBER: MI98A002792
; PRIOR FILING DATE: 1998-12-23
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: Patentin ver. 2.1
; SEQ ID NO 13
; LENGTH: 3128
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: deoD cloned
; OTHER INFORMATION: downstream ptac promoter
US-09-891-865A-13
```

```
Query Match          99.6%; Score 716.8; DB 10; Length 3128;
Best Local Similarity 99.7%; Pred. No. 5.8e-228;
Matches 718; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ATGGCTACCCACACATTAATGCAGAAATGGGCGATTTCGCTGACGTAGTTTGTATGCCA 60
DB 198 ATGGCTACCCACACATTAATGCAGAAATGGGCGATTTCGCTGACGTAGTTTGTATGCCA 257

QY 61 GGGACCCGCTGCGTGCAGAAATATATGCTGAAACTTTTCTTGAAGATGCCGCTGAAGTG 120
DB 258 GGGACCCGCTGCGTGCAGAAATATATGCTGAAACTTTTCTTGAAGATGCCGCTGAAGTG 317

QY 121 AACAAAGCTTCGCGGTATGCTGGGCTTCACCGGTACTTACAAAGGCCGCAAAATTTCCGTA 180
DB 318 AACAAAGCTTCGCGGTATGCTGGGCTTCACCGGTACTTACAAAGGCCGCAAAATTTCCGTA 377

QY 181 ATGGGTCAACGCTGTGCTGCTCCGCTGCTTCCATCTACACCAAGAACTGTATCACCAGT 240
DB 378 ATGGGTCAACGCTGTGCTGCTCCGCTGCTTCCATCTACACCAAGAACTGTATCACCAGT 437

QY 241 TTGCGGTGAAGAAATATATCCGGTGGGTTCTGTGGCGCAGTTCTGCGCGCAGTAAAA 300
DB 438 TTGCGGTGAAGAAATATATCCGGTGGGTTCTGTGGCGCAGTTCTGCGCGCAGTAAAA 497

QY 301 CTGCGGACGCTGCTATCGGTATGGGTGCTGCACCGATTCCAAAGTTAACCGCATCCGT 360
DB 498 CTGCGGACGCTGCTATCGGTATGGGTGCTGCACCGATTCCAAAGTTAACCGCATCCGT 557

QY 361 TTTAAAGACCATGACTTTGCGCGTATCGCTGACTTCGACATGGTGGGTAAACCTGTTCTCCGCTGACCTGTTTC 420
DB 558 TTTAAAGACCATGACTTTGCGCGTATCGCTGACTTCGACATGGTGGGTAAACCTGTTCTCCGCTGACCTGTTTC 617

QY 421 GCAGCTAAAGCATCTGGGTATTTGATGCTCGCGTGGGTAAACCTGTTCTCCGCTGACCTGTTTC 480
DB 618 GCAGCTAAAGCATCTGGGTATTTGATGCTCGCGTGGGTAAACCTGTTCTCCGCTGACCTGTTTC 677

QY 481 TACTCTCCGACGCGGAAATGTTTCGACGTGATCGAAATAATACGGCAATTTCTCGCGGTGAA 540
DB 678 TACTCTCCGACGCGGAAATGTTTCGACGTGATCGAAATAATACGGCAATTTCTCGCGGTGAA 737

QY 541 ATGGAAGCGGCTGGTATCTACGCGGTGCTGACGATTTGGCGCGAAGCCCTGACCATC 600
DB 738 ATGGAAGCGGCTGGTATCTACGCGGTGCTGACGATTTGGCGCGAAGCCCTGACCATC 797

QY 601 TGACCGGTATCTGACACCATCTCGCATCTCAGACGACGACCACTGCGCGTGAAGCTCAGACT 660
DB 798 TGACCGGTATCTGACACCATCTCGCATCTCAGACGACGACCACTGCGCGTGAAGCTCAGACT 857

QY 661 ACCTTCAACGACATGATCAAAATCGCACTGGAATCCGTTCTGCTGGGCGGATAAAGAGTAA 720
DB 858 ACCTTCAACGACATGATCAAAATCGCACTGGAATCCGTTCTGCTGGGCGGATAAAGAGTAA 917
```

RESULT 6

```
US-11-036-497-13
; Sequence 13 Application US/11036497
; Publication No. US20050142645A1
; GENERAL INFORMATION:
; APPLICANT: Bestetti, Giuseppina
; APPLICANT: Cali', Simona
```

```
; APPLICANT: Ghisotti, Daniela
; APPLICANT: Orsini, Gaetano
; APPLICANT: Tonon, Giancarlo
; APPLICANT: Zuffi, Gabriele
; TITLE OF INVENTION: Recombinant bacterial strains for the production of natural nucleic acid analogues thereof
; FILE REFERENCE: 02901/000J410-USO
; CURRENT APPLICATION NUMBER: US/11/036,497
; CURRENT FILING DATE: 2005-01-14
; PRIOR APPLICATION NUMBER: PCT/EP99/10416
; PRIOR FILING DATE: 1999-12-23
; PRIOR APPLICATION NUMBER: MI98A002792
; PRIOR FILING DATE: 1998-12-23
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 13
; LENGTH: 3128
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: deoD cloned downstream ptac promoter
US-11-036-497-13
```

```
Query Match          99.6%; Score 716.8; DB 24; Length 3128;
Best Local Similarity 99.7%; Pred. No. 5.8e-228;
Matches 718; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ATGGCTACCCACACATTAATGCAGAAATGGGCGATTTCGCTGACGTAGTTTGTATGCCA 60
DB 198 ATGGCTACCCACACATTAATGCAGAAATGGGCGATTTCGCTGACGTAGTTTGTATGCCA 257

QY 61 GGGACCCGCTGCGTGCAGAAATATATGCTGAAACTTTTCTTGAAGATGCCGCTGAAGTG 120
DB 258 GGGACCCGCTGCGTGCAGAAATATATGCTGAAACTTTTCTTGAAGATGCCGCTGAAGTG 317

QY 121 AACAAAGCTTCGCGGTATGCTGGGCTTCACCGGTACTTACAAAGGCCGCAAAATTTCCGTA 180
DB 318 AACAAAGCTTCGCGGTATGCTGGGCTTCACCGGTACTTACAAAGGCCGCAAAATTTCCGTA 377

QY 181 ATGGGTCAACGCTGTGCTGCTCCGCTGCTTCCATCTACACCAAGAACTGTATCACCAGT 240
DB 378 ATGGGTCAACGCTGTGCTGCTCCGCTGCTTCCATCTACACCAAGAACTGTATCACCAGT 437

QY 241 TTGCGGTGAAGAAATATATCCGGTGGGTTCTGTGGCGCAGTTCTGCGCGCAGTAAAA 300
DB 438 TTGCGGTGAAGAAATATATCCGGTGGGTTCTGTGGCGCAGTTCTGCGCGCAGTAAAA 497

QY 301 CTGCGGACGCTGCTTATCGGTATGGGTGCTGCACCGATTCCAAAGTTAACCGCATCCGT 360
DB 498 CTGCGGACGCTGCTTATCGGTATGGGTGCTGCACCGATTCCAAAGTTAACCGCATCCGT 557

QY 361 TTTAAAGACCATGACTTTGCGCGTATCGCTGACTTCGACATGGTGGGTAAACCTGTTCTCCGCTGACCTGTTTC 420
DB 558 TTTAAAGACCATGACTTTGCGCGTATCGCTGACTTCGACATGGTGGGTAAACCTGTTCTCCGCTGACCTGTTTC 617

QY 421 GCAGCTAAAGCATCTGGGTATTTGATGCTCGCGTGGGTAAACCTGTTCTCCGCTGACCTGTTTC 480
DB 618 GCAGCTAAAGCATCTGGGTATTTGATGCTCGCGTGGGTAAACCTGTTCTCCGCTGACCTGTTTC 677

QY 481 TACTCTCCGACGCGGAAATGTTTCGACGTGATCGAAATAATACGGCAATTTCTCGCGGTGAA 540
DB 678 TACTCTCCGACGCGGAAATGTTTCGACGTGATCGAAATAATACGGCAATTTCTCGCGGTGAA 737

QY 541 ATGGAAGCGGCTGGTATCTACGCGGTGCTGACGATTTGGCGCGAAGCCCTGACCATC 600
DB 738 ATGGAAGCGGCTGGTATCTACGCGGTGCTGACGATTTGGCGCGAAGCCCTGACCATC 797

QY 601 TGACCGGTATCTGACACCATCTCGCATCTCAGACGACGACCACTGCGCGTGAAGCTCAGACT 660
DB 798 TGACCGGTATCTGACACCATCTCGCATCTCAGACGACGACCACTGCGCGTGAAGCTCAGACT 857

QY 661 ACCTTCAACGACATGATCAAAATCGCACTGGAATCCGTTCTGCTGGGCGGATAAAGAGTAA 720
DB 917 ACCTTCAACGACATGATCAAAATCGCACTGGAATCCGTTCTGCTGGGCGGATAAAGAGTAA 917
```

Db 858 ACCTTCAACGACATGATCAAAATCCCAATCCGATCGGAATCCGTTCTGCTGGCGGATAAAGAGTAA 917

RESULT 7
US-09-891-865A-3
; Sequence 3, Application US/09891865A
; Publication No. US20030059670A1
; GENERAL INFORMATION:
; APPLICANT: NORPHARMA SPA
; TITLE OF INVENTION: Recombinant bacterial strains for the production of
; FILE REFERENCE: 99PC268
; CURRENT APPLICATION NUMBER: US/09/891,865A
; CURRENT FILING DATE: 2002-06-25
; PRIOR APPLICATION NUMBER: MI98A002792
; PRIOR FILING DATE: 1998-12-23
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 3383
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Plasmid
; NAME/KEY: gene
; LOCATION: (231)..(960)
; OTHER INFORMATION: deod
US-09-891-865A-3

Query Match 99.6%; Score 716.8; DB 10; Length 3383;
Best Local Similarity 99.7%; Pred. No. 6e-228;
Matches 718; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 ATGGCTACCCACACATTAATGCAGAAATGGCGGATTTTCGCTGACGTAGTTTGTATGCCA 60
Db 240 ATGGCTACCCACACATTAATGCAGAAATGGCGGATTTTCGCTGACGTAGTTTGTATGCCA 299

Qy 61 GCGGACCCGCTGCGTGCAGAAATGATTCGCTGAACTTTCCCTTGAAGATGCCCGTGAAGTG 120
Db 300 GCGGACCCGCTGCGTGCAGAAATGATTCGCTGAACTTTCCCTTGAAGATGCCCGTGAAGTG 359

Qy 121 AACAACTGTCGCGTATGCTGGCTTACCCGCTTACCAAGGCGCAAAATTTCCGTA 180
Db 360 AACAACTGTCGCGTATGCTGGCTTACCCGCTTACCAAGGCGCAAAATTTCCGTA 419

Qy 181 ATGGGTACCGTGTGGTATCCCGTCCATCTACACCAAGAACTGATCACCGAT 240
Db 420 ATGGGTACCGTGTGGTATCCCGTCCATCTACACCAAGAACTGATCACCGAT 479

Qy 241 TTCCGCGTGAAGAAATTTATCCGCGTGGGTTCTTGAAGATGCCCGTGAAGTG 300
Db 480 TTCCGCGTGAAGAAATTTATCCGCGTGGGTTCTTGAAGATGCCCGTGAAGTG 539

Qy 301 CTGCGCGACGTGTTATCGGTATCGGTGTCCTGCAACCGATTCCTCAAAAGTTTAAACCGCATCCGT 360
Db 540 CTGCGCGACGTGTTATCGGTATCGGTGTCCTGCAACCGATTCCTCAAAAGTTTAAACCGCATCCGT 599

Qy 361 TTTAAAGACCATGATCTTTCGCGCTATCGCTGACATGGTGGCGTAAACGCACTAGAT 420
Db 600 TTTAAAGACCATGATCTTTCGCGCTATCGCTGACATGGTGGCGTAAACGCACTAGAT 659

Qy 421 GCAGCTAAAGCACTGGGTATTTGATGCTCGCGTGGGTAACTGTTCTCGCGTGAACCTGTC 480
Db 660 GCAGCTAAAGCACTGGGTATTTGATGCTCGCGTGGGTAACTGTTCTCGCGTGAACCTGTC 719

Qy 481 TACTCTCCGGAACGCGAAATGTTTGAACCTGATGGAATAATACGGCATTTCTCGGCGTGAA 540
Db 720 TACTCTCCGGAACGCGAAATGTTTGAACCTGATGGAATAATACGGCATTTCTCGGCGTGAA 779

Qy 541 ATGGAAGCGGCTGTTATCTACCGCGTCCCTGAGAAATTTGGGCGGAAGCCCTGACCATC 600
Db 780 ATGGAAGCGGCTGTTATCTACCGCGTCCCTGAGAAATTTGGGCGGAAGCCCTGACCATC 839

Qy 601 TGCACCGTATCTGACCAATCCGCACTCACGAGCAGACCACTGCGGCTGAGCGTCAGACT 660
Db 840 TGCACCGTATCTGACCAATCCGCACTCACGAGCAGACCACTGCGGCTGAGCGTCAGACT 899

Qy 661 ACCTTCAACGACATGATCAAAATCCCAATCCGATCGGAATCCGTTCTGCTGGCGGATAAAGAGTAA 720
Db 900 ACCTTCAACGACATGATCAAAATCCCAATCCGATCGGAATCCGTTCTGCTGGCGGATAAAGAGTAA 959

RESULT 8
US-11-036-497-3
; Sequence 3, Application US/11036497
; Publication No. US20050142645A1
; GENERAL INFORMATION:
; APPLICANT: Bestetti, Giuseppina
; APPLICANT: Cali', Simona
; APPLICANT: Ghisotti, Daniela
; APPLICANT: Orsini, Gaetano
; APPLICANT: Tonon, Giancarlo
; APPLICANT: Zuffi, Gabriele
; TITLE OF INVENTION: Recombinant bacterial strains for the production of natural nucle
; FILE REFERENCE: 02901/000J410-US0
; CURRENT APPLICATION NUMBER: US/11/036,497
; CURRENT FILING DATE: 2005-01-14
; PRIOR APPLICATION NUMBER: PCT/EP99/10416
; PRIOR FILING DATE: 1999-12-23
; PRIOR APPLICATION NUMBER: MI98A002792
; PRIOR FILING DATE: 1998-12-23
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 3383
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: plasmid
; NAME/KEY: gene
; LOCATION: (231)..(960)
; OTHER INFORMATION: deod
US-11-036-497-3

Query Match 99.6%; Score 716.8; DB 24; Length 3383;
Best Local Similarity 99.7%; Pred. No. 6e-228;
Matches 718; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 ATGGCTACCCACACATTAATGCAGAAATGGCGGATTTTCGCTGACGTAGTTTGTATGCCA 60
Db 240 ATGGCTACCCACACATTAATGCAGAAATGGCGGATTTTCGCTGACGTAGTTTGTATGCCA 299

Qy 61 GCGGACCCGCTGCGTGCAGAAATGATTCGCTGAACTTTCCCTTGAAGATGCCCGTGAAGTG 120
Db 300 GCGGACCCGCTGCGTGCAGAAATGATTCGCTGAACTTTCCCTTGAAGATGCCCGTGAAGTG 359

Qy 121 AACAACTGTCGCGTATGCTGGCTTACCCGCTTACCAAGGCGCAAAATTTCCGTA 180
Db 360 AACAACTGTCGCGTATGCTGGCTTACCCGCTTACCAAGGCGCAAAATTTCCGTA 419

Qy 181 ATGGGTACCGTGTGGTATCCCGTCCATCTACACCAAGAACTGATCACCGAT 240
Db 420 ATGGGTACCGTGTGGTATCCCGTCCATCTACACCAAGAACTGATCACCGAT 479

Qy 241 TTCCGCGTGAAGAAATTTATCCGCGTGGGTTCTTGAAGATGCCCGTGAAGTG 300
Db 480 TTCCGCGTGAAGAAATTTATCCGCGTGGGTTCTTGAAGATGCCCGTGAAGTG 539

Qy 301 CTGCGCGACGTGTTATCGGTATCGGTGTCCTGCAACCGATTCCTCAAAAGTTTAAACCGCATCCGT 360
Db 540 CTGCGCGACGTGTTATCGGTATCGGTGTCCTGCAACCGATTCCTCAAAAGTTTAAACCGCATCCGT 599

Qy 361 TTTAAAGACCATGATCTTTCGCGCTATCGCTGACATGGTGGCGTAAACGCACTAGAT 420

```
Db 600 TTTAAAGACCATGACTTTTCGGCTATCGCTGACTTCGACATGGTGCAGTAGAT 659
QY 421 GCAGCTAAAGCAGCTGGGTATGATGCTCGCGTGGGTAAACCTGTTCTCCGCTGACCTGTTTC 480
Db 660 GCAGCTAAAGCAGCTGGGTATGATGCTCGCGTGGGTAAACCTGTTCTCCGCTGACCTGTTTC 719
QY 481 TACTCTCCGACCGCGGAAATGTTTCGACGTGATGGAATAATACGGCATTTCTCGCGGTGGA 540
Db 720 TACTCTCCGACCGCGGAAATGTTTCGACGTGATGGAATAATACGGCATTTCTCGCGGTGGA 779
QY 541 ATGGAAGCGGCTGGTATCTACGGCGTGCCTGACAGAAATTTGGCGGAAAGCCCTGACCATC 600
Db 780 ATGGAAGCGGCTGGTATCTACGGCGTGCCTGACAGAAATTTGGCGGAAAGCCCTGACCATC 839
QY 601 TGCACCGTATCTGACCAATCCGACATCTACAGACAGACCACTGCGCGTGAAGCTCAGACT 660
Db 840 TGCACCGTATCTGACCAATCCGACATCTACAGACAGACCACTGCGCGTGAAGCTCAGACT 899
QY 661 ACCTTCAACGACATGATCAAAATCGCACTGGAATCCGTTCTGCTGGCGGATAAAGAGTAA 720
Db 900 ACCTTCAACGACATGATCAAAATCGCACTGGAATCCGTTCTGCTGGCGGATAAAGAGTAA 959

RESULT 9
US-09-891-865A-14
; Sequence 14, Application US/09891865A
; Publication No. US20030059870A1
; GENERAL INFORMATION:
; APPLICANT: NORPHARMA SPA
; TITLE OF INVENTION: Recombinant bacterial strains for the production of
; FILE REFERENCE: 99DC26E
; CURRENT APPLICATION NUMBER: US/09/891,865A
; CURRENT FILING DATE: 2002-06-25
; PRIOR APPLICATION NUMBER: MI98A002792
; PRIOR FILING DATE: 1998-12-23
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 14
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: udp and deoD
; OTHER INFORMATION: cloned downstream ptac promoter
US-09-891-865A-14

Query Match 99.6%; Score 716.8; DB 10; Length 3934;
Best Local Similarity 99.7%; Pred. No. 6.5e-228;
Matches 718; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ATGGCTACCCACACATTAATGCAGAAATGGCGGATTTGCTGACGTAGTTTGTATGCCA 60
Db 198 ATGGCTACCCACACATTAATGCAGAAATGGCGGATTTGCTGACGTAGTTTGTATGCCA 257
QY 61 GCGCACCGCTGGTGGAGTATATGCTGAACCTTTCTTGAAGTGCCTGGAAGTG 120
Db 258 GCGCACCGCTGGTGGAGTATATGCTGAACCTTTCTTGAAGTGCCTGGAAGTG 317
QY 121 AACAAAGTTTCGGGTATGCTGGGCTTCAACCGGTACTTACAAAGCGCGCAAAATTTCCGTA 180
Db 318 AACAAAGTTTCGGGTATGCTGGGCTTCAACCGGTACTTACAAAGCGCGCAAAATTTCCGTA 377
QY 181 ATGGGTACCGGTATGCTGGGCTTCAACCGGTACTTACAAAGCGCGCAAAATTTCCGTA 240
Db 378 ATGGGTACCGGTATGCTGGGCTTCAACCGGTACTTACAAAGCGCGCAAAATTTCCGTA 437
QY 241 TTCCGGGTGAAGAAATTAATCCGGGTGGGTTCCTGGGCGAGTCTTCCGCGCAGCTAAA 300
Db 438 TTCCGGGTGAAGAAATTAATCCGGGTGGGTTCCTGGGCGAGTCTTCCGCGCAGCTAAA 497
QY 301 CTGCGGACGCTGCTTATCGGTATGGGTGCCTGCAACCGATTTCAAAAGTTTAACCGCATCCGT 360
```

```
Db 498 CTGCGGACGCTCGTTATCGGTATGGGTGCTGCACCGATTTCAAAGTTAACCGCATCCGT 557
QY 361 TTTAAAGACCATGACTTTTCGGCTATCGCTGACTTCGACATGGTGCAGTAGAT 420
Db 558 TTTAAAGACCATGACTTTTCGGCTATCGCTGACTTCGACATGGTGCAGTAGAT 617
QY 421 GCAGCTAAAGCAGCTGGGTATGATGCTCGCGTGGGTAAACCTGTTCTCCGCTGACCTGTTTC 480
Db 618 GCAGCTAAAGCAGCTGGGTATGATGCTCGCGTGGGTAAACCTGTTCTCCGCTGACCTGTTTC 677
QY 481 TACTCTCCGACCGCGGAAATGTTTCGACGTGATGGAATAATACGGCATTTCTCGCGGTGGA 540
Db 678 TACTCTCCGACCGCGGAAATGTTTCGACGTGATGGAATAATACGGCATTTCTCGCGGTGGA 737
QY 541 ATGGAAGCGGCTGGTATCTACGGCGTGCCTGACAGAAATTTGGCGGAAAGCCCTGACCATC 600
Db 738 ATGGAAGCGGCTGGTATCTACGGCGTGCCTGACAGAAATTTGGCGGAAAGCCCTGACCATC 797
QY 601 TGCACCGTATCTGACCAATCCGACATCTACAGACAGACCACTGCGCGTGAAGCTCAGACT 660
Db 798 TGCACCGTATCTGACCAATCCGACATCTACAGACAGACCACTGCGCGTGAAGCTCAGACT 857
QY 661 ACCTTCAACGACATGATCAAAATCGCACTGGAATCCGTTCTGCTGGCGGATAAAGAGTAA 720
Db 858 ACCTTCAACGACATGATCAAAATCGCACTGGAATCCGTTCTGCTGGCGGATAAAGAGTAA 917

RESULT 10
US-11-036-497-14
; Sequence 14, Application US/11036497
; Publication No. US20050142645A1
; GENERAL INFORMATION:
; APPLICANT: Bestetti, Giuseppina
; APPLICANT: Cali', Simona
; APPLICANT: Ghisotti, Daniela
; APPLICANT: Orsini, Gaetano
; APPLICANT: Tonon, Giancarlo
; APPLICANT: Zuffi, Gabriele
; TITLE OF INVENTION: Recombinant bacterial strains for the production of natural nucleosides and modified analogues thereof
; FILE REFERENCE: 02901/000J410-USO
; CURRENT APPLICATION NUMBER: US/11/036,497
; CURRENT FILING DATE: 2005-01-14
; PRIOR APPLICATION NUMBER: PCT/EP99/10416
; PRIOR FILING DATE: 1999-12-23
; PRIOR APPLICATION NUMBER: MI98A002792
; PRIOR FILING DATE: 1998-12-23
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 14
; LENGTH: 3934
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: udp and deoD cloned downstream ptac promoter
US-11-036-497-14

Query Match 99.6%; Score 716.8; DB 24; Length 3934;
Best Local Similarity 99.7%; Pred. No. 6.5e-228;
Matches 718; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ATGGCTACCCACACATTAATGCAGAAATGGCGGATTTGCTGACGTAGTTTGTATGCCA 60
Db 198 ATGGCTACCCACACATTAATGCAGAAATGGCGGATTTGCTGACGTAGTTTGTATGCCA 257
QY 61 GCGCACCGCTGGTGGAGTATATGCTGAACCTTTCTTGAAGTGCCTGGAAGTG 120
Db 258 GCGCACCGCTGGTGGAGTATATGCTGAACCTTTCTTGAAGTGCCTGGAAGTG 317
QY 121 AACAAAGTTTCGGGTATGCTGGGCTTCAACCGGTACTTACAAAGCGCGCAAAATTTCCGTA 180
Db 318 AACAAAGTTTCGGGTATGCTGGGCTTCAACCGGTACTTACAAAGCGCGCAAAATTTCCGTA 377
```

Qy 181 ATGGGTCAAGGTGTGGTATCCCGTCTCTGCTCCATCTACACCAAGAACTGATCACCAGT 240
Db |||||
Qy 378 ATGGGTCAAGGTGTGGTATCCCGTCTCTGCTCCATCTACACCAAGAACTGATCACCAGT 437
Db |||||
Qy 241 TTCGGCGTGAAGAAATTTATCCGGTGGGTTCCTGTGGCGAGTTCGCCGCACTGAAA 300
Db |||||
Qy 438 TTCGGCGTGAAGAAATTTATCCGGTGGGTTCCTGTGGCGAGTTCGCCGCACTGAAA 497
Db |||||
Qy 301 CTGGCGCAAGTTCGTTATCGGTATGGGTGCTGCAACCGATTCCAAAGTTAAACCGCATCGT 360
Db |||||
Qy 498 CTGGCGCAAGTTCGTTATCGGTATGGGTGCTGCAACCGATTCCAAAGTTAAACCGCATCGT 557
Db |||||
Qy 361 TTTAAAGACCATGACTTTGCGCGTATCGCTGACTTCGACATGGTGGCGTAAACGCACTAGAT 420
Db |||||
Qy 558 TTTAAAGACCATGACTTTGCGCGTATCGCTGACTTCGACATGGTGGCGTAAACGCACTAGAT 617
Db |||||
Qy 421 GCAGCTAAAGCACATGGGTATTTGATGCTGCGGTGGGTAACTGTTCTCGCGTGAACCTGTT 480
Db |||||
Qy 618 GCAGCTAAAGCACATGGGTATTTGATGCTGCGGTGGGTAACTGTTCTCGCGTGAACCTGTT 677
Db |||||
Qy 481 TACTCTCCGACGCGGAAATGTTGCACTGATGGAATAATACGGCATTTCTCGGCGTGA 540
Db |||||
Qy 678 TACTCTCCGACGCGGAAATGTTGCACTGATGGAATAATACGGCATTTCTCGGCGTGA 737
Db |||||
Qy 541 ATGAAAGCGGTGCTATCTACGGGTGCTGCTGAGAAATTTGGCGGAAAGCCCTGACCATC 600
Db |||||
Qy 738 ATGAAAGCGGTGCTATCTACGGGTGCTGCTGAGAAATTTGGCGGAAAGCCCTGACCATC 797
Db |||||
Qy 601 TGCACCGTATCTGACCATCCGCACTCAGCAGCAGACCACTGCGCGTGAACCTGAGCTGACT 660
Db |||||
Qy 798 TGCACCGTATCTGACCATCCGCACTCAGCAGCAGACCACTGCGCGTGAACCTGAGCTGACT 857
Db |||||
Qy 661 ACCTTCAACGACATGATCAAAATCGCACTGGAATCCGTTCTGCTGGGCGATAAAGAGTAA 720
Db |||||
Qy 858 ACCTTCAACGACATGATCAAAATCGCACTGGAATCCGTTCTGCTGGGCGATAAAGAGTAA 917
Db |||||

RESULT 11

US-09-891-865A-5
; Sequence 5, Application US/09891865A
; Publication No. US20030059870A1
; GENERAL INFORMATION:
; APPLICANT: NORPHARMA SPA
; TITLE OF INVENTION: Recombinant bacterial strains for the production of
; FILE REFERENCE: 99DC96E
; CURRENT APPLICATION NUMBER: US/09/891,865A
; CURRENT FILING DATE: 2002-06-25
; PRIOR APPLICATION NUMBER: MI98A002792
; PRIOR FILING DATE: 1998-12-23
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 4189
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Plasmid
; FEATURE:
; NAME/KEY: gene
; LOCATION: (243)..(1021)
; OTHER INFORMATION: udp
; FEATURE:
; NAME/KEY: gene
; LOCATION: (1037)..(1766)
; OTHER INFORMATION: deod
US-09-891-865A-5

Query Match 99.6%; Score 716.8; DB 10; Length 4189;
Best Local Similarity 99.7%; Pred. No. 6.7e-228;
Matches 718; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 ATGGCTACCCACACATTAATGAGAAATGGCGGATTTTCGCTGAGTAGTTTGTATGCCA 60
Db |||||
Qy 1046 ATGGCTACCCACACATTAATGAGAAATGGCGGATTTTCGCTGAGTAGTTTGTATGCCA 1105
Db |||||
Qy 61 GGCGACCCGCTGCGTGGGAGATATTTGCTGAAACTTTCTTTGAAGATGCCCGTGAAGTG 120
Db |||||
Qy 1106 GGCGACCCGCTGCGTGGGAGATATTTGCTGAAACTTTCTTTGAAGATGCCCGTGAAGTG 1165
Db |||||
Qy 121 AACCAACGTTCCGCGTATGCTGGGCTTCACCGGTATCTTCAAAAGCCGCAAAATTTCCGTA 180
Db |||||
Qy 1166 AACCAACGTTCCGCGTATGCTGGGCTTCACCGGTATCTTCAAAAGCCGCAAAATTTCCGTA 1225
Db |||||
Qy 181 ATGGGTCAAGGTGTGGTATCCCGTCTGCTCCATCTACACCAAGAACTGATCACCAGT 240
Db |||||
Qy 1226 ATGGGTCAAGGTATGGGTATCCCGTCTGCTCCATCTACACCAAGAACTGATCACCAGT 1285
Db |||||
Qy 241 TTGGCGGTGAAGAAATTTATCCGGTGGGTTCCTGTGGCGAGTTCGTGCCGACGTAATA 300
Db |||||
Qy 1286 TTGGCGGTGAAGAAATTTATCCGGTGGGTTCCTGTGGCGAGTTCGTGCCGACGTAATA 1345
Db |||||
Qy 301 CTGGCGCAAGTTCGTTATCGGTATGGGTGCTGCAACCGATTCCAAAGTTAAACCGCATCGT 360
Db |||||
Qy 1346 CTGGCGCAAGTTCGTTATCGGTATGGGTGCTGCAACCGATTCCAAAGTTAAACCGCATCGT 1405
Db |||||
Qy 361 TTTAAAGACCATGACTTTGCGCGTATCGCTGACTTCGACATGGTGGCGTAAACGCACTAGAT 420
Db |||||
Qy 1406 TTTAAAGACCATGACTTTGCGCGTATCGCTGACTTCGACATGGTGGCGTAAACGCACTAGAT 1465
Db |||||
Qy 421 GCAGCTAAAGCACATGGGTATTTGATGCTGCGGTGGGTAACTGTTCTCGCGTGAACCTGTT 480
Db |||||
Qy 1466 GCAGCTAAAGCACATGGGTATTTGATGCTGCGGTGGGTAACTGTTCTCGCGTGAACCTGTT 1525
Db |||||
Qy 481 TACTCTCCGACGCGGAAATGTTTCGAGTGTATGGAATAATACGGCATTTCTCGGCGTGA 540
Db |||||
Qy 1526 TACTCTCCGACGCGGAAATGTTTCGAGTGTATGGAATAATACGGCATTTCTCGGCGTGA 1585
Db |||||
Qy 541 ATGAAAGCGGTGCTATCTACGGGTGCTGCTGAGAAATTTGGCGGAAAGCCCTGACCATC 600
Db |||||
Qy 1586 ATGAAAGCGGTGCTATCTACGGGTGCTGCTGAGAAATTTGGCGGAAAGCCCTGACCATC 1645
Db |||||
Qy 601 TGCACCGTATCTGACCATCCGCACTCAGCAGCAGACCACTGCGCGTGAACCTGAGCTGACT 660
Db |||||
Qy 1646 TGCACCGTATCTGACCATCCGCACTCAGCAGCAGACCACTGCGCGTGAACCTGAGCTGACT 1705
Db |||||
Qy 661 ACCTTCAACGACATGATCAAAATCGCACTGGAATCCGTTCTGCTGGGCGATAAAGAGTAA 720
Db |||||
Qy 1706 ACCTTCAACGACATGATCAAAATCGCACTGGAATCCGTTCTGCTGGGCGATAAAGAGTAA 1765
Db |||||

RESULT 12

US-11-036-497-5
; Sequence 5, Application US/11036497
; Publication No. US20050142645A1
; GENERAL INFORMATION:
; APPLICANT: Bestetti, Giuseppina
; APPLICANT: Cali', Simona
; APPLICANT: Ghisotti, Daniela
; APPLICANT: Orsini, Gaetano
; APPLICANT: Tonon, Giancarlo
; APPLICANT: Zuffi, Gabriele
; TITLE OF INVENTION: Recombinant bacterial strains for the production of natural nucle
; FILE REFERENCE: 02901/000J410-USO
; CURRENT APPLICATION NUMBER: US/11/036,497
; CURRENT FILING DATE: 2005-01-14
; PRIOR APPLICATION NUMBER: PCT/EP99/10416
; PRIOR FILING DATE: 1999-12-23
; PRIOR APPLICATION NUMBER: MI98A002792
; PRIOR FILING DATE: 1998-12-23
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5
; LENGTH: 4189

; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: plasmid
; FEATURE:
; NAME/KEY: gene
; LOCATION: (243)..(1021)
; OTHER INFORMATION: udp
; FEATURE:
; NAME/KEY: gene
; LOCATION: (1037)..(1766)
; OTHER INFORMATION: deod
US-11-036-497-5

Query Match 99.6%; Score 716.8; DB 24; Length 4189;
Best Local Similarity 99.7%; Pred. No. 6.7e-228;
Matches 718; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 ATGGCTACCCACACATTAATGCAGAAATGGCGATTTCCGCTGACGTAGTTTGTATGCCA 60
DB 1046 ATGGCTACCCACACATTAATGCAGAAATGGCGATTTCCGCTGACGTAGTTTGTATGCCA 1105
QY 61 GGGACCCGCTGCGTGCAGATATATGCTGAACTTTCTTGAAGATGCCCGTGAAGTG 120
DB 1106 GGGACCCGCTGCGTGCAGATATATGCTGAACTTTCTTGAAGATGCCCGTGAAGTG 1165
QY 121 AACACGTTCCGCGTATGCTGGGCTTACCGGTACTTACAAAGCCGCAAAATTTCCGTA 180
DB 1166 AACACGTTCCGCGTATGCTGGGCTTACCGGTACTTACAAAGCCGCAAAATTTCCGTA 1225
QY 181 ATGGGTCAACGGTGTGGTATCCCGTCTCGCTGCATCTACACCAAGAACTGATCACCGAT 240
DB 1226 ATGGGTCAACGGTATGGGTATCCCGTCTCGCTGCATCTACACCAAGAACTGATCACCGAT 1285
QY 241 TTCCGCGTGAAGAAATTAATCCGCGTGGGTTCCTGTGGCGCAGTTCTGCGCGACGTAATA 300
DB 1286 TTCCGCGTGAAGAAATTAATCCGCGTGGGTTCCTGTGGCGCAGTTCTGCGCGACGTAATA 1345
QY 301 CTGGCGACGCTGTTATCGGTATGGGTGCTGACCGATTTCCAAAGTTAACCGCATCCGT 360
DB 1346 CTGGCGACGCTGTTATCGGTATGGGTGCTGACCGATTTCCAAAGTTAACCGCATCCGT 1405
QY 361 TTTAAAGACCATGACTTTGCGGCTATCGCTGACTTTCGACATGGTGGCGTAACGCGATAGAT 420
DB 1406 TTTAAAGACCATGACTTTGCGGCTATCGCTGACTTTCGACATGGTGGCGTAACGCGATAGAT 1465
QY 421 GCAGCTAAAGCAGCTGGGTATGATGCTCGCGTGGGTAACTGTTCTCCGCTGACCTGTTTC 480
DB 1466 GCAGCTAAAGCAGCTGGGTATGATGCTCGCGTGGGTAACTGTTCTCCGCTGACCTGTTTC 1525
QY 481 TACTCTCCGACCGCGAAATGTTCCGAGCTGATGGAAATACGGCAATTTCTCGCGGTGGAA 540
DB 1526 TACTCTCCGACCGCGAAATGTTCCGAGCTGATGGAAATACGGCAATTTCTCGCGGTGGAA 1585
QY 541 ATGGAAAGCGCTGGTATCTACGGCGTCCGCTGCAGAAATTTGGCGCGAAAGCCCTGACCATC 600
DB 1586 ATGGAAAGCGCTGGTATCTACGGCGTCCGCTGCAGAAATTTGGCGCGAAAGCCCTGACCATC 1645
QY 601 TGCAACCGTATCTGACCAATCCGCACTCAACGAGCAGACCACTGCGCGTGAAGCTGACACT 660
DB 1646 TGCAACCGTATCTGACCAATCCGCACTCAACGAGCAGACCACTGCGCGTGAAGCTGACACT 1705
QY 661 ACCTTCAACGACATGATCAAAATCGCACTGGAATCCGTTCTGCTGGCGGATAAAGAGTAA 720
DB 1706 ACCTTCAACGACATGATCAAAATCGCACTGGAATCCGTTCTGCTGGCGGATAAAGAGTAA 1765

RESULT 13
US-10-198-034-5
; Sequence 5, Application US/10198034
; Publication No. US20030077269A1
; GENERAL INFORMATION:
; APPLICANT: Sorscher, Eric J.

; APPLICANT: Bennett, Leonard L.
; APPLICANT: Parker, William B.
; APPLICANT: Waud, William
; APPLICANT: Gadi, Vijaykrishna K.
; TITLE OF INVENTION: RECOMBINANT BACTERIAL CELLS FOR DELIVERY OF PNP TO TUMOR CELLS
; FILE REFERENCE: UAB-12407/22
; CURRENT APPLICATION NUMBER: US/10/198,034
; CURRENT FILING DATE: 2002-07-18
; PRIOR APPLICATION NUMBER: 08/702,181
; PRIOR FILING DATE: 1996-08-23
; PRIOR APPLICATION NUMBER: 08/122,321
; PRIOR FILING DATE: 1993-09-14
; PRIOR APPLICATION NUMBER: 60/064,676
; PRIOR FILING DATE: 1997-10-31
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5
; LENGTH: 5013
; TYPE: DNA
; ORGANISM: Escherichia coli
US-10-198-034-5

Query Match 99.6%; Score 716.8; DB 14; Length 5013;
Best Local Similarity 99.7%; Pred. No. 7.2e-228;
Matches 718; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 ATGGCTACCCACACATTAATGCAGAAATGGCGATTTCCGCTGACGTAGTTTGTATGCCA 60
DB 413 ATGGCTACCCACACATTAATGCAGAAATGGCGATTTCCGCTGACGTAGTTTGTATGCCA 472
QY 61 GGGACCCGCTGCGTGCAGATATATGCTGAACTTTCTTGAAGATGCCCGTGAAGTG 120
DB 473 GGGACCCGCTGCGTGCAGATATATGCTGAACTTTCTTGAAGATGCCCGTGAAGTG 532
QY 121 AACACGTTCCGCGTATGCTGGGCTTACCGGTACTTACAAAGCCGCAAAATTTCCGTA 180
DB 533 AACACGTTCCGCGTATGCTGGGCTTACCGGTACTTACAAAGCCGCAAAATTTCCGTA 592
QY 181 ATGGGTCAACGGTGTGGTATCCCGTCTCGCTGCATCTACACCAAGAACTGATCACCGAT 240
DB 593 ATGGGTCAACGGTATGGGTATCCCGTCTCGCTGCATCTACACCAAGAACTGATCACCGAT 652
QY 241 TTCCGCGTGAAGAAATTAATCCGCGTGGGTTCCTGTGGCGCAGTTCTGCGCGACGTAATA 300
DB 653 TTCCGCGTGAAGAAATTAATCCGCGTGGGTTCCTGTGGCGCAGTTCTGCGCGACGTAATA 712
QY 301 CTGGCGACGCTGTTATCGGTATGGGTGCTGACCGATTTCCAAAGTTAACCGCATCCGT 360
DB 713 CTGGCGACGCTGTTATCGGTATGGGTGCTGACCGATTTCCAAAGTTAACCGCATCCGT 772
QY 361 TTTAAAGACCATGACTTTGCGGCTATCGCTGACTTTCGACATGGTGGCGTAACGCGATAGAT 420
DB 773 TTTAAAGACCATGACTTTGCGGCTATCGCTGACTTTCGACATGGTGGCGTAACGCGATAGAT 832
QY 421 GCAGCTAAAGCAGCTGGGTATGATGCTCGCGTGGGTAACTGTTCTCCGCTGACCTGTTTC 480
DB 833 GCAGCTAAAGCAGCTGGGTATGATGCTCGCGTGGGTAACTGTTCTCCGCTGACCTGTTTC 892
QY 481 TACTCTCCGACCGCGAAATGTTTCGAGCTGATGGAAATACGGCAATTTCTCGCGGTGGAA 540
DB 893 TACTCTCCGACCGCGAAATGTTTCGAGCTGATGGAAATACGGCAATTTCTCGCGGTGGAA 952
QY 541 ATGGAAAGCGCTGGTATCTACGGCGTCCGCTGCAGAAATTTGGCGCGAAAGCCCTGACCATC 600
DB 953 ATGGAAAGCGCTGGTATCTACGGCGTCCGCTGCAGAAATTTGGCGCGAAAGCCCTGACCATC 1012
QY 601 TGCAACCGTATCTGACCAATCCGCACTCAACGAGCAGACCACTGCGCGTGAAGCTGACACT 660
DB 1013 TGCAACCGTATCTGACCAATCCGCACTCAACGAGCAGACCACTGCGCGTGAAGCTGACACT 1072
QY 661 ACCTTCAACGACATGATCAAAATCGCACTGGAATCCGTTCTGCTGGCGGATAAAGAGTAA 720
DB 1073 ACCTTCAACGACATGATCAAAATCGCACTGGAATCCGTTCTGCTGGCGGATAAAGAGTAA 1132

Qy	421	GCAGCTAAAGCACTGGGTATTGATGCTCGCGTGGGTAAACCTGTTCTCCGCTGACCTGTTTC	480
Db	1742	GCAGCTAAAGCACTGGGTATTGATGCTCGCGTGGGTAAACCTGTTCTCCGCTGACCTGTTTC	1801
Qy	481	TACTCTCCGGACCGCGAAATGTTTCGACGTGATGGAAAAATACGGCAATTCTCGGCGTGGAA	540
Db	1802	TACTCTCCGGACCGCGAAATGTTTCGACGTGATGGAAAAATACGGCAATTCTCGGCGTGGAA	1861
Qy	541	ATGGAAAGCGGCTGGTATCTACGGGTCGCTGCAGAAATTTGGCGGAAAGCCCTGACCATC	600
Db	1862	ATGGAAAGCGGCTGGTATCTACGGGTCGCTGCAGAAATTTGGCGGAAAGCCCTGACCATC	1921
Qy	601	TGCACCGTATCTGACCAACATCCGCACCTCACGAGCAGACCACTGCCGCTGAGCGTCAGACT	660
Db	1922	TGCACCGTATCTGACCAACATCCGCACCTCACGAGCAGACCACTGCCGCTGAGCGTCAGACT	1981
Qy	661	ACCTTCAAACGACATGATCAAAAATCGCACTGGAATCCGTTCTGTGGGCGATAAAGAGTAA	720
Db	1982	ACCTTCAAACGACATGATCAAAAATCGCACTGGAATCCGTTCTGTGGGCGATAAAGAGTAA	2041

Search completed: July 23, 2005, 01:38:14
Job time : 806 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 22, 2005, 17:40:37 ; Search time 4167 Seconds
(without alignments)
6576.980 Million cell updates/sec

Title: US-10-035-300A-1

Perfect score: 720

Sequence: 1 atggctaccacacattaa.....tgctggcgataaagagtaa 720

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:

- 1: gb_est1.*
- 2: gb_est2.*
- 3: gb_hic.*
- 4: gb_est3.*
- 5: gb_est4.*
- 6: gb_est5.*
- 7: gb_est6.*
- 8: gb_gse1.*
- 9: gb_gse2.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	272.6	37.9	767	7	CV205284
C 2	272.6	37.9	769	7	CV204676
C 3	272.6	37.9	771	7	CV219418
C 4	272.6	37.9	792	7	CV204675
5	272.6	37.9	794	7	CV205285
6	272.6	37.9	794	7	CV219419
7	260.4	36.2	721	7	CV205285
C 8	191.6	26.6	1589	8	BH770690
C 9	169	23.5	169	7	CF306721
10	169	23.5	169	7	CF306860
11	150	20.8	161	7	CF307418
12	121.8	16.9	921	8	BH138974
C 13	110	15.3	863	8	AZ680406
C 14	101.8	14.1	861	8	BH166036
15	101	14.0	864	8	AZ675191
C 16	100	13.9	878	8	AZ533877
C 17	93.8	13.0	865	8	AZ669801
C 18	80.2	11.1	226	7	CV213810
19	80.2	11.1	250	7	CV213811
C 20	77	10.7	882	8	AZ681446
21	74	10.3	879	8	AZ539664
C 22	66	9.2	861	8	AZ669012
C 23	65	9.0	715	9	AG266877
24	64.6	9.0	234	7	CV203815

C 25	64.6	9.0	234	7	CV203816	CV203816	EST863526
C 26	59.4	8.2	904	8	AZ536366	AZ536366	ENTCX45TF
C 27	57	7.9	801	9	CC924665	t078f14ba	CC924665
C 28	56.4	7.8	898	8	BH148646	ENTPE80TR	BH148646
C 29	46.8	6.5	890	8	AZ529376	ENTCV88TF	AZ529376
30	43	6.0	742	4	BM170790	EST573313	BM170790
31	42	5.8	572	4	BM165171	EST567694	BM165171
32	41.4	5.8	602	7	CN896963	CN896963	010630AAZ
33	41.4	5.8	645	4	BM170428	BM170428	EST572951
34	41.4	5.8	690	4	BM162943	EST565466	BM162943
35	41.4	5.8	709	4	BM160915	EST563438	BM160915
36	41.4	5.8	734	4	BM159186	EST561709	BM159186
37	41.4	5.8	768	4	BM159582	EST562105	BM159582
38	40	5.6	706	6	C96538	C96538	010630AAZ
39	39.8	5.5	260	7	CO866368	CO866368	010630AAZ
40	39.8	5.5	543	7	CN918728	CN918728	030207ABP
41	39.8	5.5	559	7	CN903541	CN903541	021018ABC
42	39.8	5.5	604	4	BM162201	BM162201	EST564724
43	39.8	5.5	649	7	CN902206	CN902206	021014ABC
44	39.8	5.5	681	7	CN903108	CN903108	021016ABC
45	39.8	5.5	743	7	CO866081	CO866081	Mdb5014f

ALIGNMENTS

RESULT 1
CV205284/c
LOCUS CV205284 767 bp mRNA linear EST 16-SEP-2004
DEFINITION EST864994 non-normalized T1 cDNA library Trichomonas vaginalis cDNA
clone rvtTAC60 3' end, mRNA sequence.
ACCESSION CV205284
VERSION CV205284.1 GI:52152264
KEYWORDS EST.
SOURCE Trichomonas vaginalis
ORGANISM Trichomonas vaginalis
Eukaryota; Parabasalidea; Trichomonadida; Trichomonadidae; Trichomonadinae; Trichomonas.

REFERENCE 1 (bases 1 to 767)
AUTHORS Carlton,J.M., Dyall,S., Johnson,P.J. and Fraser,C.M.
TITLE The complete genome sequence of the sexually transmitted parasite Trichomonas vaginalis
JOURNAL Unpublished (2004)
COMMENT Other ESTs: EST864995
Contact: Jane Carlton
Parasite Genomics Group
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-530-9319
Fax: 301-838-0208
Email: carlton@tigr.org
Seq primer: polydT 24 base.

FEATURES

source
1..767
/organism="Trichomonas vaginalis"
/mol_type="mRNA"
/strain="T1"
/db_xref="taxon:5722"
/clone="TWTAC60"
/clone_lib="non-normalized T1 cDNA library"
/note="Vector: Lambda Triplex2; Site_1: SfiI; Site_2: SfiI; T. vaginalis strain T1 library constructed from cDNA, made in lambda Triplex2. Inserts cloned unidirectionally in the SfiI and SfiI sites. Mass excision of library produced inserts in pTriplex2 plasmid. Inserts sequenced from both 5' and 3' ends using Triplex2 sequencing primer and polydT 24 bp primer respectively."

ORIGIN

Query Match 37.9%; Score 272.6; DB 7; Length 767;
Best Local Similarity 62.0%; Pred. No. 3e-73;
Matches 431; Conservative 0; Mismatches 264; Indels 0; Gaps 0;


```
CV219418/c
LOCUS       CV219418      771 bp      mRNA      linear      EST 16-SEP-2004
DEFINITION   EST879128 non-normalized T1 cDNA library Trichomonas vaginalis cDNA
clone TVTX26 3' end, mRNA sequence.
ACCESSION   CV219418
VERSION     CV219418.1
KEYWORDS    Trichomonas vaginalis
SOURCE      Trichomonas vaginalis
ORGANISM    Trichomonas vaginalis
REFERENCE   1 (bases 1 to 771)
AUTHORS     Carlton,J.M., Dyall,S., Johnson,P.J. and Fraser,C.M.
TITLE       The complete genome sequence of the sexually transmitted parasite Trichomonas vaginalis
JOURNAL     Unpublished (2004)
COMMENT     Other ESTs: EST879129
            Contact: Jane Carlton
            Parasite Genomics Group
            The Institute for Genomic Research
            9712 Medical Center Drive, Rockville, MD 20850, USA
            Tel: 301-530-9319
            Fax: 301-838-0208
            Email: carlton@tigr.org
            Seq primer: polydT 24 base.
            Location/Qualifiers
                1..771
                /organism="Trichomonas vaginalis"
                /mol_type="mRNA"
                /strain="T1"
                /db_xref="taxon:5722"
                /clone="TVTX26"
                /note="Vector: Lambda Triplex2; Site 1: SfiA; Site 2: SfiB; T. vaginalis strain T1 library constructed from cDNA, made in lambda Triplex2. Inserts cloned unidirectionally in the SfiA and SfiB sites. Mass excision of library produced inserts in pTriplex2 plasmid. Inserts sequenced from both 5' and 3' ends using Triplex2 sequencing primer and polydT 24 bp primer respectively."

FEATURES             source
    Query Match       37.9%; Score 272.6; DB 7; Length 771;
    Best Local Similarity 62.0%; Pred. No. 3e-73;
    Matches 431; Conservative 0; Mismatches 264; Indels 0; Gaps 0;

    QY 1 ATGGCTACCCACACATTAATGCGAAGTATATTCGTAAGCTTTCCTTGAAGATGCCCGTGAAGTG 60
    Db 1 ATGGCTACCCACACATTAATGCGAAGTATATTCGTAAGCTTTCCTTGAAGATGCCCGTGAAGTG 60

    QY 61 GCGGACCCGCTGCGTGCAGAAATATATTCGTAAGCTTTCCTTGAAGATGCCCGTGAAGTG 120
    Db 61 GCGGACCCGCTGCGTGCAGAAATATATTCGTAAGCTTTCCTTGAAGATGCCCGTGAAGTG 120

    QY 699 GCGGATCACTCCGTCGCAAGCTCATTCGTAAGCTTTCCTTGAAGATGCCCGTGAAGTG 640
    Db 699 GCGGATCACTCCGTCGCAAGCTCATTCGTAAGCTTTCCTTGAAGATGCCCGTGAAGTG 640

    QY 121 AACACAGTTTCGCGGTATCTCGGCTTCAACGGTACTTACAAAGGCCGCAAAATTTCCGTA 180
    Db 121 AACACAGTTTCGCGGTATCTCGGCTTCAACGGTACTTACAAAGGCCGCAAAATTTCCGTA 180

    QY 639 AACTCTGTTCTGGAATGCTGTTTACAGGCACATACAAAGGCCAACCCTCTCTGTT 580
    Db 639 AACTCTGTTCTGGAATGCTGTTTACAGGCACATACAAAGGCCAACCCTCTCTGTT 580

    QY 181 ATGGGTCAAGTGTGGTATCCCGTCTGCTCCATCTACCAAGAACTGATCAGCAT 240
    Db 181 ATGGGTCAAGTGTGGTATCCCGTCTGCTCCATCTACCAAGAACTGATCAGCAT 240

    QY 579 ATGGGTCAAGCAATGGGTATCCCATTTCCATCTACGCTGAAGAACTCTACAAAGCTT 520
    Db 579 ATGGGTCAAGCAATGGGTATCCCATTTCCATCTACGCTGAAGAACTCTACAAAGCTT 520

    QY 241 TTCGCGGTGAAGAAATATATCCGCTGGGTTCCTGTGGCGAGTTCTCGCGCACGTAATA 300
    Db 241 TTCGCGGTGAAGAAATATATCCGCTGGGTTCCTGTGGCGAGTTCTCGCGCACGTAATA 300

    QY 519 TACAAGGTTAAGACCATCATCCGTCGCGCACTTTCGCGCACAGTTGATCCAAATGTTTCA 460
    Db 519 TACAAGGTTAAGACCATCATCCGTCGCGCACTTTCGCGCACAGTTGATCCAAATGTTTCA 460

    QY 301 CTGGCGCAAGTGTGTTATCGGTATGCGGTGCGTGCACCGATTTCAAAGTTTAAACCGCATCCGT 360
    Db 301 CTGGCGCAAGTGTGTTATCGGTATGCGGTGCGTGCACCGATTTCAAAGTTTAAACCGCATCCGT 360

    QY 459 GTCCGCGCATGTCTGCAATGTCACAGCATCTGCGCACAGATTCAAATGTTTAAACAGATGGC 400
    Db 459 GTCCGCGCATGTCTGCAATGTCACAGCATCTGCGCACAGATTCAAATGTTTAAACAGATGGC 400

    QY 361 TTTAAAGACCATGACTTTTGGCGCTATCCGTCGATTCGACATGTTGCGTAAACGCGATAGAT 420
    Db 361 TTTAAAGACCATGACTTTTGGCGCTATCCGTCGATTCGACATGTTGCGTAAACGCGATAGAT 420

    QY 399 CTTCCTGGCCACGATTTCCCGAGCTACAGCCAACTTCGAAAGTCGTTAGTGCCTTAGTTGAA 340
    Db 399 CTTCCTGGCCACGATTTCCCGAGCTACAGCCAACTTCGAAAGTCGTTAGTGCCTTAGTTGAA 340

CV204675
LOCUS       CV204675      792 bp      mRNA      linear      EST 16-SEP-2004
DEFINITION   EST864385 non-normalized T1 cDNA library Trichomonas vaginalis cDNA
clone TVTA905 3' end, mRNA sequence.
ACCESSION   CV204675
VERSION     CV204675.1
KEYWORDS    Trichomonas vaginalis
SOURCE      Trichomonas vaginalis
ORGANISM    Trichomonas vaginalis
REFERENCE   1 (bases 1 to 792)
AUTHORS     Carlton,J.M., Dyall,S., Johnson,P.J. and Fraser,C.M.
TITLE       The complete genome sequence of the sexually transmitted parasite Trichomonas vaginalis
JOURNAL     Unpublished (2004)
COMMENT     Other ESTs: EST864386
            Contact: Jane Carlton
            Parasite Genomics Group
            The Institute for Genomic Research
            9712 Medical Center Drive, Rockville, MD 20850, USA
            Tel: 301-530-9319
            Fax: 301-838-0208
            Email: carlton@tigr.org
            Seq primer: polydT 24 base.
            Location/Qualifiers
                1..792
                /organism="Trichomonas vaginalis"
                /mol_type="mRNA"
                /strain="T1"
                /db_xref="taxon:5722"
                /clone="TVTA905"
                /note="Vector: Lambda Triplex2; Site 1: SfiA; Site 2: SfiB; T. vaginalis strain T1 library constructed from cDNA, made in lambda Triplex2. Inserts cloned unidirectionally in the SfiA and SfiB sites. Mass excision of library produced inserts in pTriplex2 plasmid. Inserts sequenced from both 5' and 3' ends using Triplex2 sequencing primer and polydT 24 bp primer respectively."

    QY 421 GCAGCTAAAGCAGCTGGGTATTGATGCTCGCGTGGGTAAACCTGTTCTCGCTGACCTGTTTC 480
    Db 421 GCAGCTAAAGCAGCTGGGTATTGATGCTCGCGTGGGTAAACCTGTTCTCGCTGACCTGTTTC 480

    QY 339 TCGGCCAAGGCTCTCAACATCCCAACACAGGTGTTGAAAGGCCCTACTCAACAGATATCTTC 280
    Db 339 TCGGCCAAGGCTCTCAACATCCCAACACAGGTGTTGAAAGGCCCTACTCAACAGATATCTTC 280

    QY 481 TACTCTCCGAGCGGCGAAATGTTGACGTGATGGAATAATAGCGCATTTCTCGCGGTGGAA 540
    Db 481 TACTCTCCGAGCGGCGAAATGTTGACGTGATGGAATAATAGCGCATTTCTCGCGGTGGAA 540

    QY 279 TACAGCAAGGAACAAGGATTAAATGAGGCCCTTGTCTAGTACCACTTCAATTCTGTCGAA 220
    Db 279 TACAGCAAGGAACAAGGATTAAATGAGGCCCTTGTCTAGTACCACTTCAATTCTGTCGAA 220

    QY 541 ATGGAAGCGGTGGTATTCTACGGCTGCTCAGAAATTTGGCGGAAAGCCCTGACCATC 600
    Db 541 ATGGAAGCGGTGGTATTCTACGGCTGCTCAGAAATTTGGCGGAAAGCCCTGACCATC 600

    QY 219 ATGGAATCTGCTGGTCCCTTCCAAATGCTGATTAATACTATGGCGCAAGAGCTGGCTGCATC 160
    Db 219 ATGGAATCTGCTGGTCCCTTCCAAATGCTGATTAATACTATGGCGCAAGAGCTGGCTGCATC 160

    QY 601 TGCACCGTATCTGACCAATCGCATCTCAGCAGCAGACCACTGCGCGCTGAGCGTCAGACT 660
    Db 601 TGCACCGTATCTGACCAATCGCATCTCAGCAGCAGACCACTGCGCGCTGAGCGTCAGACT 660

    QY 159 TGCACAGTTTCGGATCATCATCATCACACAGTCGCTACACCAAGAGAGCGCCAGACA 100
    Db 159 TGCACAGTTTCGGATCATCATCATCACACAGTCGCTACACCAAGAGAGCGCCAGACA 100

    QY 661 ACCTTCAAGCAGCATGATCAAAATGCGCATCGGAATC 695
    Db 661 ACCTTCAAGCAGCATGATCAAAATGCGCATCGGAATC 695

    QY 99 TCATTCCAGAAACATGATCAAGATCGCTCTTGAAGC 65
    Db 99 TCATTCCAGAAACATGATCAAGATCGCTCTTGAAGC 65

RESULT 4
CV204675
LOCUS       CV204675      792 bp      mRNA      linear      EST 16-SEP-2004
DEFINITION   EST864385 non-normalized T1 cDNA library Trichomonas vaginalis cDNA
clone TVTA905 3' end, mRNA sequence.
ACCESSION   CV204675
VERSION     CV204675.1
KEYWORDS    Trichomonas vaginalis
SOURCE      Trichomonas vaginalis
ORGANISM    Trichomonas vaginalis
REFERENCE   1 (bases 1 to 792)
AUTHORS     Carlton,J.M., Dyall,S., Johnson,P.J. and Fraser,C.M.
TITLE       The complete genome sequence of the sexually transmitted parasite Trichomonas vaginalis
JOURNAL     Unpublished (2004)
COMMENT     Other ESTs: EST864386
            Contact: Jane Carlton
            Parasite Genomics Group
            The Institute for Genomic Research
            9712 Medical Center Drive, Rockville, MD 20850, USA
            Tel: 301-530-9319
            Fax: 301-838-0208
            Email: carlton@tigr.org
            Seq primer: polydT 24 base.
            Location/Qualifiers
                1..792
                /organism="Trichomonas vaginalis"
                /mol_type="mRNA"
                /strain="T1"
                /db_xref="taxon:5722"
                /clone="TVTA905"
                /note="Vector: Lambda Triplex2; Site 1: SfiA; Site 2: SfiB; T. vaginalis strain T1 library constructed from cDNA, made in lambda Triplex2. Inserts cloned unidirectionally in the SfiA and SfiB sites. Mass excision of library produced inserts in pTriplex2 plasmid. Inserts sequenced from both 5' and 3' ends using Triplex2 sequencing primer and polydT 24 bp primer respectively."

    QY 1 ATGGCTACCCACACATTAATGCGAAGTATATTCGTAAGCTTTCCTTGAAGATGCCCGTGAAGTG 60
    Db 1 ATGGCTACCCACACATTAATGCGAAGTATATTCGTAAGCTTTCCTTGAAGATGCCCGTGAAGTG 60

    QY 61 GCGGACCCGCTGCGTGCAGAAATATATTCGTAAGCTTTCCTTGAAGATGCCCGTGAAGTG 120
    Db 61 GCGGACCCGCTGCGTGCAGAAATATATTCGTAAGCTTTCCTTGAAGATGCCCGTGAAGTG 120

    QY 699 GCGGATCACTCCGTCGCAAGCTCATTCGTAAGCTTTCCTTGAAGATGCCCGTGAAGTG 640
    Db 699 GCGGATCACTCCGTCGCAAGCTCATTCGTAAGCTTTCCTTGAAGATGCCCGTGAAGTG 640

    QY 121 AACACAGTTTCGCGGTATCTCGGCTTCAACGGTACTTACAAAGGCCGCAAAATTTCCGTA 180
    Db 121 AACACAGTTTCGCGGTATCTCGGCTTCAACGGTACTTACAAAGGCCGCAAAATTTCCGTA 180

    QY 639 AACTCTGTTCTGGAATGCTGTTTACAGGCACATACAAAGGCCAACCCTCTCTGTT 580
    Db 639 AACTCTGTTCTGGAATGCTGTTTACAGGCACATACAAAGGCCAACCCTCTCTGTT 580

    QY 181 ATGGGTCAAGTGTGGTATCCCGTCTGCTCCATCTACCAAGAACTGATCAGCAT 240
    Db 181 ATGGGTCAAGTGTGGTATCCCGTCTGCTCCATCTACCAAGAACTGATCAGCAT 240

    QY 579 ATGGGTCAAGCAATGGGTATCCCATTTCCATCTACGCTGAAGAACTCTACAAAGCTT 520
    Db 579 ATGGGTCAAGCAATGGGTATCCCATTTCCATCTACGCTGAAGAACTCTACAAAGCTT 520

    QY 241 TTCGCGGTGAAGAAATATATCCGCTGGGTTCCTGTGGCGAGTTCTCGCGCACGTAATA 300
    Db 241 TTCGCGGTGAAGAAATATATCCGCTGGGTTCCTGTGGCGAGTTCTCGCGCACGTAATA 300

    QY 519 TACAAGGTTAAGACCATCATCCGTCGCGCACTTTCGCGCACAGTTGATCCAAATGTTTCA 460
    Db 519 TACAAGGTTAAGACCATCATCCGTCGCGCACTTTCGCGCACAGTTGATCCAAATGTTTCA 460

    QY 301 CTGGCGCAAGTGTGTTATCGGTATGCGGTGCGTGCACCGATTTCAAAGTTTAAACCGCATCCGT 360
    Db 301 CTGGCGCAAGTGTGTTATCGGTATGCGGTGCGTGCACCGATTTCAAAGTTTAAACCGCATCCGT 360

    QY 459 GTCCGCGCATGTCTGCAATGTCACAGCATCTGCGCACAGATTCAAATGTTTAAACAGATGGC 400
    Db 459 GTCCGCGCATGTCTGCAATGTCACAGCATCTGCGCACAGATTCAAATGTTTAAACAGATGGC 400

    QY 361 TTTAAAGACCATGACTTTTGGCGCTATCCGTCGATTCGACATGTTGCGTAAACGCGATAGAT 420
    Db 361 TTTAAAGACCATGACTTTTGGCGCTATCCGTCGATTCGACATGTTGCGTAAACGCGATAGAT 420

    QY 399 CTTCCTGGCCACGATTTCCCGAGCTACAGCCAACTTCGAAAGTCGTTAGTGCCTTAGTTGAA 340
    Db 399 CTTCCTGGCCACGATTTCCCGAGCTACAGCCAACTTCGAAAGTCGTTAGTGCCTTAGTTGAA 340
```



```

ACCESSION   CV219419.1  GI:52166399
VERSION      EST.
KEYWORDS     Trichomonas vaginalis
SOURCE       Trichomonas vaginalis
ORGANISM     Eukaryota; Parabasalidea; Trichomonada; Trichomonadida;
              Trichomonadidae; Trichomonadinae; Trichomonas.
REFERENCE    1 (bases 1 to 794)
AUTHORS      Carlton, J.M., Dyall, S., Johnson, P.J. and Fraser, C.M.
TITLE        The complete genome sequence of the sexually transmitted parasite
              Trichomonas vaginalis
JOURNAL      Unpublished (2004)
COMMENT      Other_ESTs: EST879128
              Contact: Jane Carlton
              Parasite Genomics Group
              The Institute for Genomic Research
              9712 Medical Center Drive, Rockville, MD 20850, USA
              Tel: 301-530-9319
              Fax: 301-838-0208
              Email: carlton@tigr.org
              Seq primer: lambda Triplex2.
              Location/Qualifiers
FEATURES             1..794
                     /organism="Trichomonas vaginalis"
                     /mol_type="mRNA"
                     /strain="T1"
                     /db_xref="taxon:5722"
                     /clone="TVTCX26"
                     /clone_lib="non-normalized T1 cDNA library"
                     /note="Vector: Lambda triplex2; Site_1: SfiI; Site_2:
                     SfiB; T. vaginalis strain T1 library constructed from
                     cDNA, made in lambda Triplex2. Inserts cloned
                     unidirectionally in the SfiI and SfiB sites. Mass excision
                     of library produced inserts in pTriplex2 plasmid. Inserts
                     sequenced from both 5' and 3' ends using Triplex2
                     sequencing primer and polydr 24 bp primer respectively."
ORIGIN
Query Match      37.9%; Score 272.6; DB 7; Length 794;
Best Local Similarity 62.0%; Pred. No. 3.1e-73;
Matches 431; Conservative 0; Mismatches 264; Indels 0; Gaps 0;

Qy 1 ATGGCTACCCACACATTAAATGCGAATAATGGCGGATTCGCTGACGTAGTTGATGCCA 60
Db |||||
Qy 13 ATGGCAACACCCCATAACTCTGCAAGGTTGGCGATTTTCGCTGACAGATTTCTATGTGC 72
Db |||||
Qy 61 GGCGACCGCTGCGTGGCAAGTATATGCTGAACTTTCTTGAAGATGCCCGTGAAGTG 120
Db |||||
Qy 73 GGCGATCCACTCCGTCGCAAGCTCATTTCTGATTAATCTTGAGAACGCTAAGCAAGTC 132
Db |||||
Qy 121 AACACGTTTCGGGTATGCTGGGCTTCAACCGGTACTTACAAGGCGCAAAATTTCCGTA 180
Db |||||
Qy 133 AACTCTGTTCTGGNATGCTTGGTTTACAGGCACATACAGGGCAACACCTCTCTGTT 192
Db |||||
Qy 181 ATGGGTACCGGTGTTGGTATCCCGTCTGCTCCATCTACACCAAGAACTGATCACCAGT 240
Db |||||
Qy 193 ATGGGTATGTCATGGCATGGGTATCCCATCCATTTCCATCTACGCTGAGAACTCTACAACGTT 252
Db |||||
Qy 241 TTGGCGGTGAGAAATTTATCCGGTGGGTTCTGTGGCGGAGTTCTGCCGACGTAAAA 300
Db |||||
Qy 253 TACAAGGTTAAGACCATCATCCGTGTCGGCACTTTCGGGCAAGTTGATCCAAATGTTCCAC 312
Db |||||
Qy 301 CTGGCGGACGTCGTTATTCGGTATGGGTGCTGCACCGGATTCAAAGTTAAACGGATCCGT 360
Db |||||
Qy 313 GTCCGGNATGTCGATTTGTACAGCATCTCGCACAGATTTCAATGTTAAGAGATGGCG 372
Db |||||
Qy 361 TTTAAAGACCAAGACTTTGCGGCTATGCTGATGCTGCAATGGTGGCGTAAACGCAAGTAGAT 420
Db |||||
Qy 373 CTTCTTGGCCACGATTTCCAGCTACAGCCAACTTCGAAGTCGTAGTGGCTTAGTGA 432
Db |||||
Qy 421 CGAGTAAAGCACTGGGTATTGATGCTCGGTGGGTAACTGTTCTCGCTGACCTGTC 480
Db |||||
Qy 433 TCCGCCAAGGCTCTCAACATCCCAACACACAGGTTGGAAAGGCGCTACTCAACAGATATCTTC 492

```

```

Qy 481 TACTCTCCGACGCGGAAATGTTTCGACGTGATGGAATAATACGGCATTTCTCGGCGTGAA 540
Db |||||
Qy 493 TACAGCAAGGAACAAGGATTAAATGAGGCCCTTCTCAGTACCACTTCTATTGCTGTCGA 552
Db |||||
Qy 541 ATGGAAGCGGCTGTTATCTACGGCGTTCGTCAGAAATTTGGCGGAAGCCCTGACCATC 600
Db |||||
Qy 553 ATGGAATCTGCTGCTGCTCCCTTCCCAATTTGCTGATAACTATGCGCAAGAGCTGGCTGCATC 612
Db |||||
Qy 601 TGCACCGTATCTGACCAACATCCGCACTCACCAGCAGACCACTGCGGCTGAGCGTCAGACT 660
Db |||||
Qy 613 TGCACAGTTTCGATCATCATCATCACACAGGTCGCGTACACAGAGAGCGCCAGACA 672
Db |||||
Qy 661 ACCTTCAACGACATGATCAAAATCGCACTGGAATC 695
Db |||||
Qy 673 TCATTCAGAACATGATCAAGATCGCTTTGAAGC 707
Db |||||

RESULT 7
CO577885
LOCUS      TVEST085809 Tv30236_PT cDNA Library Trichomonas vaginalis cDNA 5',
DEFINITION mRNA sequence.
ACCESSION CO577885
VERSION    CO577885.1 GI:50408335
KEYWORDS   EST.
SOURCE     Trichomonas vaginalis
ORGANISM   Eukaryota; Parabasalidea; Trichomonada; Trichomonadida;
              Trichomonadidae; Trichomonadinae; Trichomonas.
REFERENCE  1 (bases 1 to 721)
AUTHORS    Zhou, Y., Shu, W.M., Huang, S.C.C., Huang, K.Y. and Tang, P.
TITLE      Analysis of Gene Expression Profile in Trichomonas vaginalis by EST
              Sequencing
JOURNAL     Unpublished (2003)
COMMENT     Contact: Tang, P.
              Molecular Regulation and Bioinformatics Laboratory, College of
              Medicine
              Chang Gung University
              259 Wenhsia 1st. Road, Kweishan, Taoyuan 333, Taiwan
              Tel: +886 3 3283016 EXT5136
              Fax: +886 3 3283031
              Email: petang@mail.cgu.edu.tw
              PCR Primers
              FORWARD: T7
              BACKWARD: T3
              Seq primer: T3.
              Location/Qualifiers
FEATURES             1..721
                     /organism="Trichomonas vaginalis"
                     /mol_type="mRNA"
                     /db_xref="taxon:5722"
                     /cell_line="ATCC30236"
                     /dev_stage="Trophozoites at mid-log phase"
                     /lab_host="XLI Blue-MRF"
                     /clone_lib="TV30236_PT cDNA Library"
                     /note="Vector: Lambda ZAP-Express (Stratagene); Site_1:
                     EcoRI; Site_2: XhoI"
ORIGIN
Query Match      36.2%; Score 260.4; DB 7; Length 721;
Best Local Similarity 61.8%; Pred. No. 1.9e-69;
Matches 414; Conservative 0; Mismatches 256; Indels 0; Gaps 0;

Qy 28 ATGGGCGATTTTCGCTGACGTAGTTTTCATGCCAGCGCACCGCTCGTCCGAAGTATATT 87
Db |||||
Qy 41 AGGGGCGATTTTCGCTGAAACAGTCTCATGTGCGGTGATCCACTCCGCGCTAAGCTCAAT 100
Db |||||
Qy 88 GCTCAAACTTTCTTGAAGATGCCGTGAAGTGAACCAAGTTCGCGGTATGCTGGGCTTC 147
Db |||||
Qy 101 GCTGAGACATATCTTGAANAATCCAAAGCTTGTCAACATATGTTCTGGGATTCAGAGGTAC 160
Db |||||
Qy 148 ACCGGTACTTACAAAGGCGCAAAATTTCCGTAATGGGTACAGGTGTTGGTATCCCGTCC 207

```

[illegible]

RESULT 8	
BH770690/c	
LOCUS	BH770690
DEFINITION	LLMGtag441 MG1363 Random Sequence Tag Library Lactococcus lactis subsp. cremoris genomic, genomic survey sequence.
	1589 bp DNA linear GSS 01-MAY-2002

REFERENCE
1 (bases 1 to 1589)
Lactococcus.
AUTHORS
Bolotin, A., Ehrlich, S.D. and Sorokin, A.
TITLE
Studies of genomes of dairy bacteria *Lactococcus lactis*
JOURNAL
Sci. Aliments (2002) In press
COMMENT
Contact: Sorokin A

		/db xref=taxon:1359"					
		/clone lib="MG1363 Random Sequence Tag Library"					
		/notes=vector: pSGM2; Site 1: SmaI; Library of					
		chromosomal fragments of L.lactis strain MG1363 was					
		prepared by partial AluI digestion or by sonication."					
 ORIGIN							
	Query Match	26.6%	Score 191.6;	DB 8;	Length 1589;		
	Best Local Similarity	58.1%	Pred. No. 6.3e-48;				
	Matches 360;	Conservative 0;	Mismatches 254;	Indels 6;	Gaps 1;		
Qy	1	ATGGCTACCCACACATTAAATGCAGAAATCGGCGCATTTTCGCTCAGCTAGTATTTTGATGCCA	60				
Dd	646	ATGCCAACACCACATATCGAAGCTCAAAAAGGGGAAATCGCAGATAAAATTCCTTACCA	587				
Qy	61	GGGACCCGTGCGTGGCAGTAGTATATGCTGGAACCTTTCCCTTCAAGATGCCCGTGAAGTG	120				
Dd	586	GGAGATCCACTTCGCGCAAAATTTATGCGAGAAACCTTCCTTGAAGATGCTGTGCAATTT	527				
Qy	121	AACAACGTTCCGCGTATGCTGGGCTTCACCGGTACTTTACAAAGCCGCCAAAAATTTCCGTA	180				
Dd	526	AACNAGTTTCGCGGAATGCTTGCTTTACTTGGAACCTTACAAGGCCCATCGTGTTCGTGA	467				
Qy	181	ATGGGTCAACGGTGTGTGATCCCGTCTGCTCCATCTACACAAAGAATCGATCACCGAT	240				
Dd	466	ATGGGAACAGGAATGGGAAATTCCTTCAATCTCAATCTATGCTAATGAATTGATTACAGAA	407				
Qy	241	TTGGCGGTGAAGAAAATTAATCCGGTGGGTTCTGTGGCGCAGTTCGTCCGCGCAGTAAAA	300				
Dd	406	TATGGCGTTAAAAAGATTAAATTCGTGTGGGACTGCTGGTTTTCTGTAATGAAGATGTTTCAT	347				
Qy	301	CTGGCGCAGCTCGTATPCGGTATGGGTGCTGCACCGATTCCAAAGTTAACCCGCATCCGT	360				
Dd	346	ATTGCGACCTTGTCATTTGTTGAGGCGACGACAACTTCAGCCAATGTCCTCGTCATGAT	287				
Qy	361	TTTAAAGACCATGACTTTGCGCGTATCGCTGCACTTCGACATGTGTGGGTAAACGACAGTAGAT	420				
Dd	286	TTTCTCTGATTTGATTTCCCACAAAATTTGCTGATTTTGTGATTTGCTTGATAAGGCTTATCAT	227				
Qy	421	GCAGCTAAAGCACTGGGTATTTGATGCTCGGTGGGTAAACCTGTTCTCCGCTGACCTGTTTC	480				
Dd	226	ATTGCAAAAAGACCTTGGAAATTACAACATCACGTGTGGGAATATTTCTAGTCTGATTTGTTTC	167				
Qy	481	TACTCTCCGGACGCGGAAATGTTTCGACGTGATGGAATAATACGCATTTCTCGCGGTGAAA	540				
Dd	166	TATGGTGGACGAGATGCAGTTAAAG-----TTGGAANAATCGGTGTGAAGGCTGTTCGAA	113				
Qy	541	ATGGAACGGCTGGTATCTACGGCGTCCGTGCAGAAATTTGGCGCGAAAGCCCTGCACCATC	600				
Dd	112	ATGGAACGAGCTGGTCTTTATATCTTTGGGGCAAAATATAAGTGCACAACTTTGGGAATT	53				
Qy	601	TGCACCGTATCTGACCAAT	620				
Dd	52	ATGACAATTTCTGACCATAT	33				

RESULT 9	CF306721	169 bp	mRNA	linear	EST 15-AUG-2003
LOCUS	HDAL--04-K17.g1	OSHDAC1-overexpressing transgenic rice lambda phage			
DEFINITION	cDNA library I (HDAL) <i>Oryza sativa</i> (japonica cultivar-group) cDNA clone HDAL--04-K17, mRNA sequence.				
ACCESSION	CF306721				
VERSION	CF306721.1	GI:33678482			
KEYWORDS	EST.				
SOURCE	<i>Oryza sativa</i> (japonica cultivar-group)				
ORGANISM	<i>Oryza sativa</i> (japonica cultivar-group)				
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzaceae; <i>Oryza</i> .				
REFERENCE	1 (bases 1 to 169)				
AUTHORS	Kim, J. S., Jun, K. M., Cheong, P. J., Kim, M. J., Lee, T. H., Shin, Y. C., Song, S. I., Kim, J. K., Kim, Y. K. and Nahm, B. H.				

TITLE Large-scale Sequencing Analysis of Rice ESTs
JOURNAL Unpublished (2003)
COMMENT Contact: Nahm B.H.
Genomics and Genetics Institute, GreenGene Biotech Inc.; Division
of BioScience and Bioinformatics, Myongui University
Yongin, Kyeonggi, Korea
Tel: 82 31 330 6193
Fax: 82 31 321 6355
Email: bhnahm@gbio.com, bhnahm@bio.myongji.ac.kr.

FEATURES
source
1. .169
/organism="Oryza sativa (japonica cultivar-group)"
/mol_type="mRNA"
/cultivar="Nackdong"
/db_xref="taxon:39947"
/clone="HDA1--04-K17"
/tissue_type="callus"
/dev_stage="proliferated callus on 2N6 media for 2 weeks"
/lab_host="E.coli SOLR"
/clone_lib="OSHDA1-overexpressing transgenic rice lambda
phage cDNA library I (HDA1)"
/note="Vector: pBluescript SK(+); Site 1: EcoRI; Site 2:
XhoI; Callus was treated with ABA(20um) for 1hour. cDNA
was inserted into lambda Uni-ZAP XR vector at 5' end with
EcoRI and 3' end with XhoI site. mRNA was derived from
rice Histone Deacetylase overexpression line."

ORIGIN
Query Match 23.5%; Score 169; DB 7; Length 169;
Best Local Similarity 100.0%; Pred. No. 3.4e-41;
Matches 169; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 277 GGCGCAGTCTCGCCGACGTAACAACTGCGCGACGTCGTTATCGGTATGGTGCGCTGCAC 336
Db 1 GGCGCAGTCTCGCCGACGTAACAACTGCGCGACGTCGTTATCGGTATGGTGCGCTGCAC 60
Qy 337 GATTCCAAAGTTAACCGCATCCGTTTTAAAGACCATGACTTTGCCGCTATCGCTGACTTC 396
Db 61 GATTCCAAAGTTAACCGCATCCGTTTTAAAGACCATGACTTTGCCGCTATCGCTGACTTC 120
Qy 397 GACATGGTGCGTAAACGAGTAGATGACGTAAGCACTGGGTATTGATG 445
Db 121 GACATGGTGCGTAAACGAGTAGATGACGTAAGCACTGGGTATTGATG 169

RESULT 10
LOCUS CF306860
DEFINITION HDAL1--05-A22.g1 OshDA1-overexpressing transgenic rice lambda phage
clone library I (HDA1) Oryza sativa (japonica cultivar-group) cDNA
CF306860 169 bp mRNA linear EST 15-AUG-2003
ACCESSION
VERSION CF306860.1 GI:33678621
KEYWORDS
SOURCE Oryza sativa (japonica cultivar-group)
ORGANISM Oryza sativa (japonica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzae; Oryza.
1 (Bases 1 to 169)
Kim, J.S., Jun, K.M., Cheong, P.J., Kim, M.J., Lee, T.H., Shin, Y.C.,
Song, S.I., Kim, J.K., Kim, Y.-K. and Nahm, B.H.
Large-scale Sequencing Analysis of Rice ESTs
Unpublished (2003)
Contact: Nahm B.H.
Genomics and Genetics Institute, GreenGene Biotech Inc.; Division
of BioScience and Bioinformatics, Myongui University
Yongin, Kyeonggi, Korea
Tel: 82 31 330 6193
Fax: 82 31 321 6355
Email: bhnahm@gbio.com, bhnahm@bio.myongji.ac.kr.

FEATURES
source
1. .169
/organism="Oryza sativa (japonica cultivar-group)"
/mol_type="mRNA"
/cultivar="Nackdong"
/db_xref="taxon:39947"
/clone="HDA1--05-A22"
/tissue_type="callus"
/dev_stage="proliferated callus on 2N6 media for 2 weeks"
/lab_host="E.coli SOLR"
/clone_lib="OSHDA1-overexpressing transgenic rice lambda
phage cDNA library I (HDA1)"
/note="Vector: pBluescript SK(+); Site 1: EcoRI; Site 2:
XhoI; Callus was treated with ABA(20um) for 1hour. cDNA
was inserted into lambda Uni-ZAP XR vector at 5' end with
EcoRI and 3' end with XhoI site. mRNA was derived from
rice Histone Deacetylase overexpression line."

ORIGIN
Query Match 23.5%; Score 169; DB 7; Length 169;
Best Local Similarity 100.0%; Pred. No. 3.4e-41;
Matches 169; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 277 GGCGCAGTCTCGCCGACGTAACAACTGCGCGACGTCGTTATCGGTATGGTGCGCTGCAC 336
Db 1 GGCGCAGTCTCGCCGACGTAACAACTGCGCGACGTCGTTATCGGTATGGTGCGCTGCAC 60
Qy 337 GATTCCAAAGTTAACCGCATCCGTTTTAAAGACCATGACTTTGCCGCTATCGCTGACTTC 396
Db 61 GATTCCAAAGTTAACCGCATCCGTTTTAAAGACCATGACTTTGCCGCTATCGCTGACTTC 120
Qy 397 GACATGGTGCGTAAACGAGTAGATGACGTAAGCACTGGGTATTGATG 445
Db 121 GACATGGTGCGTAAACGAGTAGATGACGTAAGCACTGGGTATTGATG 169

RESULT 11
LOCUS CF307418
DEFINITION HDAL1--06-K08.g1 OshDA1-overexpressing transgenic rice lambda phage
clone library I (HDA1) Oryza sativa (japonica cultivar-group) cDNA
CF307418 161 bp mRNA linear EST 15-AUG-2003
ACCESSION
VERSION CF307418.1 GI:33679179
KEYWORDS
SOURCE Oryza sativa (japonica cultivar-group)
ORGANISM Oryza sativa (japonica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzae; Oryza.
1 (Bases 1 to 161)
Kim, J.S., Jun, K.M., Cheong, P.J., Kim, M.J., Lee, T.H., Shin, Y.C.,
Song, S.I., Kim, J.K., Kim, Y.-K. and Nahm, B.H.
Large-scale Sequencing Analysis of Rice ESTs
Unpublished (2003)
Contact: Nahm B.H.
Genomics and Genetics Institute, GreenGene Biotech Inc.; Division
of BioScience and Bioinformatics, Myongui University
Yongin, Kyeonggi, Korea
Tel: 82 31 330 6193
Fax: 82 31 321 6355
Email: bhnahm@gbio.com, bhnahm@bio.myongji.ac.kr.

FEATURES
source
1. .161
/organism="Oryza sativa (japonica cultivar-group)"
/mol_type="mRNA"
/cultivar="Nackdong"
/db_xref="taxon:39947"
/clone="HDA1--06-K08"
/tissue_type="callus"
/dev_stage="proliferated callus on 2N6 media for 2 weeks"
/lab_host="E.coli SOLR"
/clone_lib="OSHDA1-overexpressing transgenic rice lambda
phage cDNA library I (HDA1)"
/note="Vector: pBluescript SK(+); Site 1: EcoRI; Site 2:
XhoI; Callus was treated with ABA(20um) for 1hour. cDNA
was inserted into lambda Uni-ZAP XR vector at 5' end with
EcoRI and 3' end with XhoI site. mRNA was derived from
rice Histone Deacetylase overexpression line."

/organism="Oryza sativa (japonica cultivar-group)"
/mol_type="mRNA"
/cultivar="Nackdong"
/db_xref="taxon:39947"
/clone="HDA1--05-A22"
/tissue_type="callus"
/dev_stage="proliferated callus on 2N6 media for 2 weeks"
/lab_host="E.coli SOLR"
/clone_lib="OSHDA1-overexpressing transgenic rice lambda
phage cDNA library I (HDA1)"
/note="Vector: pBluescript SK(+); Site 1: EcoRI; Site 2:
XhoI; Callus was treated with ABA(20um) for 1hour. cDNA
was inserted into lambda Uni-ZAP XR vector at 5' end with
EcoRI and 3' end with XhoI site. mRNA was derived from
rice Histone Deacetylase overexpression line."

ORIGIN
Query Match 23.5%; Score 169; DB 7; Length 169;
Best Local Similarity 100.0%; Pred. No. 3.4e-41;
Matches 169; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 277 GGCGCAGTCTCGCCGACGTAACAACTGCGCGACGTCGTTATCGGTATGGTGCGCTGCAC 336
Db 1 GGCGCAGTCTCGCCGACGTAACAACTGCGCGACGTCGTTATCGGTATGGTGCGCTGCAC 60
Qy 337 GATTCCAAAGTTAACCGCATCCGTTTTAAAGACCATGACTTTGCCGCTATCGCTGACTTC 396
Db 61 GATTCCAAAGTTAACCGCATCCGTTTTAAAGACCATGACTTTGCCGCTATCGCTGACTTC 120
Qy 397 GACATGGTGCGTAAACGAGTAGATGACGTAAGCACTGGGTATTGATG 445
Db 121 GACATGGTGCGTAAACGAGTAGATGACGTAAGCACTGGGTATTGATG 169

RESULT 11
LOCUS CF307418
DEFINITION HDAL1--06-K08.g1 OshDA1-overexpressing transgenic rice lambda phage
clone library I (HDA1) Oryza sativa (japonica cultivar-group) cDNA
CF307418 161 bp mRNA linear EST 15-AUG-2003
ACCESSION
VERSION CF307418.1 GI:33679179
KEYWORDS
SOURCE Oryza sativa (japonica cultivar-group)
ORGANISM Oryza sativa (japonica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzae; Oryza.
1 (Bases 1 to 161)
Kim, J.S., Jun, K.M., Cheong, P.J., Kim, M.J., Lee, T.H., Shin, Y.C.,
Song, S.I., Kim, J.K., Kim, Y.-K. and Nahm, B.H.
Large-scale Sequencing Analysis of Rice ESTs
Unpublished (2003)
Contact: Nahm B.H.
Genomics and Genetics Institute, GreenGene Biotech Inc.; Division
of BioScience and Bioinformatics, Myongui University
Yongin, Kyeonggi, Korea
Tel: 82 31 330 6193
Fax: 82 31 321 6355
Email: bhnahm@gbio.com, bhnahm@bio.myongji.ac.kr.

FEATURES
source
1. .161
/organism="Oryza sativa (japonica cultivar-group)"
/mol_type="mRNA"
/cultivar="Nackdong"
/db_xref="taxon:39947"
/clone="HDA1--06-K08"
/tissue_type="callus"
/dev_stage="proliferated callus on 2N6 media for 2 weeks"
/lab_host="E.coli SOLR"
/clone_lib="OSHDA1-overexpressing transgenic rice lambda
phage cDNA library I (HDA1)"
/note="Vector: pBluescript SK(+); Site 1: EcoRI; Site 2:
XhoI; Callus was treated with ABA(20um) for 1hour. cDNA
was inserted into lambda Uni-ZAP XR vector at 5' end with
EcoRI and 3' end with XhoI site. mRNA was derived from
rice Histone Deacetylase overexpression line."

XhoI; Callus was treated with ABA(20um) for 1hour. cDNA was inserted into lambda Uni-ZAP XR vector at 5' end with EcoRI and 3' end with XhoI site. mRNA was derived from rice Histone Deacetylase overexpression line."

ORIGIN

Query Match 20.8%; Score 150; DB 7; Length 161;
Best Local Similarity 99.4%; Pred. No. 2.8e-35;
Matches 161; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

Qy 284 TTTCGCCGACGTAACAACTGCGGACGTCGTTATCGGTATGGTGCCTGCACCGGATCCA 343
Db 1 TTTCGCCGACGTAACAACTGCGGACGTCGTTATCGGTATGGTGCCTGCACCGGATCCA 60

Qy 344 AAGTTAACCGCATCCGTTTAAAGACCATGACTTTGGCGGTATCGCTGACATTCGACATGG 403
Db 61 AAGTTAACCGCATCCGTTTAAAGA-CATGACTTTGGCGGTATCGCTGACATTCGACATGG 119

Qy 404 TGCCTAACGCGTAGATGCGAGCTAAAGCAGCTGGGTATTGATG 445
Db 120 TGCCTAACGCGTAGATGCGAGCTAAAGCAGCTGGGTATTGATG 161

RESULT 12

BH138974 921 bp DNA linear GSS 07-AUG-2001
LOCUS ENT0V94TF Entamoeba histolytica Sheared DNA Entamoeba histolytica
DEFINITION genomic, genomic survey sequence.

ACCESSION

BH138974

KEYWORDS

GSS.

SOURCE

Entamoeba histolytica

ORGANISM

Eukaryota; Entamoebidae; Entamoeba.

REFERENCE

1 (bases 1 to 921)

AUTHORS

Loftus, B., Wang, Z., Van Aken, S. and Fraser, C.

TITLE

Determination of clone end sequences from Entamoeba histolytica

JOURNAL

HM1:IMSS sheared DNA library (2001)

COMMENT

Unpublished (2001)
Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Fax: 301 838 3543
Email: bjloftus@tigr.org
Clones are derived from the Entamoeba histolytica HM1:IMSS sheared DNA library

Seq primer: M13-Forward

Class: shotgun

High quality sequence start: 2

High quality sequence stop: 563.

FEATURES

source

1. .921
/organism="Entamoeba histolytica"
/mol_type="genomic DNA"
/strain="HM1:IMSS"
/db_xref="taxon:5759"
/clone_lib="Entamoeba histolytica Sheared DNA"
/note="Vector: pHOS1; Site 1: Bst I; Constructed at The Institute for Genomic Research (TIGR), Rockville, MD.
Genomic DNA isolated from broth cultures of E. histolytica using a method described by Clark and Diamond (Clark, C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a method for isolate identification. Exp. Parasitol. 77:450.). The DNA was mechanically sheared to give a tight size distribution (~2 kb). The v + i method used for the library construction is described in detail in Smith, H.O. and Venter, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. In Genome Sequencing: A Practical Approach, eds. M. Vaudin and B. Barrel, Oxford University Press, 1999)."

ORIGIN

Query Match 16.9%; Score 121.8; DB 8; Length 921;
Best Local Similarity 53.2%; Pred. No. 2.9e-26;
Matches 304; Conservative 0; Mismatches 262; Indels 5; Gaps 2;

Qy 5 CTACCCACACATTAATGCAGAAATGGCGATTTCGCTGACGTAGTGTTCGATGCCAGCG 64
Db 128 CAACTCCACATAATGGTCTAAATACGGAGAGATTGCTGAACTGCTTATGGCTGGT 187

Qy 65 ACCCGCTGCGTGGAGATATATTCTGCTGAAACTTTCTTGAAGATGCCCGGAGAGTGAACA 124
Db 188 ATCCATTAAGAGTAAACCTTCTTGCAGACACTTATTTCGACTGATGTTGTTCAATATAA 247

Qy 125 ACGTTCGCGTATGCTGGGCTTCACCGGTACTTACAAAGGCCCAAAATTTTCGTTATGG 184
Db 248 GTGTTAGAGAGCAGTAGGATATACCTGATATTAAGAGGAGTGAAATTAAGTTCAAG 307

Qy 185 GTCACGCTGTGTGATCCCGTCTCTGCTCCATCTACACCAAGAACTGATCACCGATTTTCG 244
Db 308 CTCATGCTGCTGATGCTCCATCTATTGGAATTTATGCATATGAATTAATTTTCTATG 367

Qy 245 GCGTGAAGAAAATATCCCGCTGGGTTCTGCTGGCGCAGTTCTGCCGACGTAAGAACTGC 304
Db 368 GAGTAAAGAGAAATATTAGAAATTTGTTCTGCTGGAGCTTTTGATGAAAGTTTAAACTTG 427

Qy 305 GCGACGTCGTTATCGGTATGGGTGCTGACCGCATTTCCAAAGTTTAAACCGCATCCGTTT 364
Db 428 GAGATATGTTATGTAATGGAGCATGTTATGATTTCCAAATTTTGAAGCAATATGATA 487

Qy 365 AAGACCATGACTTTGCGCTATCGCTGATTCGACATGTCGTAACGTAACGCAAGTAGATGCG 424
Db 488 TTCAGGTAATACTCATGTTATGCTGATTTCCAACTTTGCAGAGAAGCAGTTGATGTCAG 547

Qy 425 CTAAGCAGCTGGTATGATGCTCGGTGGTAACTGTTCTCGCTGACCTGTTCTACT 484
Db 548 CTGAAAACCTGGATATAGAT-ATAAAGTAAGGAATATTATTCTGCTAATTAATTTCTATG 606

Qy 485 CTCGGACGCGCAATGTTTCGAGCTGATGGAATAATACGGCATTTCTCGGCGTGGAATGG 544
Db 607 ATGATGAGATCAC-TCGGAGCATGGAAGATGGGAGTACTTGTCTGTAGAAATG 662

Qy 545 AAGCGGCTGGTATCTACGGCGTCGCTGCAGA 575
Db 663 AAGCAGCTGCTCTTTATATGATTGCAGAGA 693

RESULT 13

AZ680406/c

LOCUS

DEFINITION

AZ680406

ACCESSION

AZ680406

VERSION

AZ680406.1

KEYWORDS

GSS.

SOURCE

Entamoeba histolytica

ORGANISM

Entamoeba histolytica

REFERENCE

1 (bases 1 to 863)

AUTHORS

Loftus, B., Van Aken, S. and Fraser, C.

TITLE

Determination of clone end sequences from Entamoeba histolytica

JOURNAL

HM1:IMSS sheared DNA library

COMMENT

Unpublished (2000)

Contact: Brendan J Loftus

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0208

Fax: 301 838 3543

Email: bjloftus@tigr.org

Clones are derived from the Entamoeba histolytica HM1:IMSS sheared DNA library

Seq primer: M13-Forward

Class: shotgun

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 18, 2005, 21:26:10 ; Search time 43 Seconds
(without alignments)
414.910 Million cell updates/sec

Title: US-10-035-300A-2
Perfect score: 1225
Sequence: 1 MATPHINAEAGDFADVLMPCDPLRAKYIAETFLDAREVNNVRGMLGFTGTGKGRKISV 239

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues
Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA.*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pap.*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pap.*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pap.*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pap.*
5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pap.*
6: /cgn2_6/ptodata/1/iaa/backfiles.pap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1162	94.9	243	4	US-09-489-039A-8802
2	1057	86.3	264	4	US-09-543-681A-6543
3	653	53.3	221	4	US-09-902-540-10882
4	607	49.6	244	4	US-09-107-532A-4839
5	591	48.2	237	4	US-09-134-000C-6013
6	576.5	47.1	236	4	US-09-583-110-2698
7	576.5	47.1	237	4	US-09-107-433-4570
8	572	46.7	243	3	US-09-134-001C-4239
9	538.5	44.0	227	4	US-09-634-238-214
10	245.5	20.0	298	4	US-09-489-039A-8031
11	227.5	18.6	265	4	US-09-543-681A-5709
12	139	11.3	236	4	US-09-107-532A-6816
13	136	11.1	238	4	US-09-489-039A-12428
14	131	10.7	242	4	US-09-543-681A-6066
15	130	10.6	232	3	US-09-347-878-18
16	130	10.6	232	4	US-09-711-164-397
17	122.5	10.0	236	4	US-09-134-000C-5104
18	116.5	9.5	308	4	US-09-328-352-4660
19	115	9.4	228	4	US-09-710-279-2936
20	115	9.4	228	4	US-09-710-279-3196
21	115	9.4	233	3	US-09-134-001C-3643
22	108	8.8	230	4	US-09-583-110-5212
23	108	8.8	239	4	US-09-107-433-3429
24	106	8.7	254	4	US-09-583-110-5019
25	103.5	8.4	330	4	US-09-949-016-11585
26	101	8.2	237	4	US-09-107-433-4653
27	91.5	7.5	435	4	US-09-248-796A-17067

28	89.5	7.3	390	4	US-09-452-937A-41	Sequence 41, Appl
29	89	7.3	218	4	US-09-134-000C-4009	Sequence 4009, Ap
30	89	7.3	639	4	US-09-252-991A-18903	Sequence 18903, A
31	87	7.1	830	4	US-09-252-991A-20619	Sequence 20619, A
32	85	6.9	607	4	US-09-540-236-2006	Sequence 2006, Ap
33	84.5	6.9	389	4	US-09-489-039A-14234	Sequence 14234, A
34	84.5	6.9	389	4	US-09-949-016-9488	Sequence 9488, Ap
35	84	6.9	546	4	US-09-396-154-34	Sequence 34, Appl
36	81.5	6.7	445	1	US-08-674-168-30	Sequence 30, Appl
37	81.5	6.7	445	3	US-08-985-908-20	Sequence 20, Appl
38	81.5	6.7	445	3	US-08-852-730-5	Sequence 5, Appl
39	81.5	6.7	445	3	US-08-985-916-12	Sequence 12, Appl
40	80.5	6.6	390	4	US-09-452-937A-20	Sequence 20, Appl
41	80	6.5	249	4	US-09-107-532A-6722	Sequence 6722, Ap
42	79.5	6.5	363	4	US-09-358-321C-12	Sequence 12, Appl
43	79.5	6.5	393	4	US-09-358-321C-6	Sequence 6, Appl
44	79.5	6.5	411	4	US-09-543-681A-7404	Sequence 7404, Ap
45	79.5	6.5	713	4	US-09-270-767-45101	Sequence 45101, A

ALIGNMENTS

RESULT 1
US-09-489-039A-8802
; Sequence 8802, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; PRIOR FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 8802
; LENGTH: 243
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-8802

Query Match	94.9%	Score 1162;	DB 4;	Length 243;
Best Local Similarity	94.6%	Pred. No. 8.6e-128;		
Matches 226;	Conservative	6;	Mismatches 7;	Indels 0; Gaps 0;
QY	1	MATPHINAEAGDFADVLMPCDPLRAKYIAETFLDAREVNNVRGMLGFTGTGKGRKISV	60	
Db	5	MATPHINAEAGDFADVLMPCDPLRAKYIAETFLDAREVNNVRGMLGFTGTGKGRKISV	64	
QY	61	MGHGVGPSCTSIYTKELITDFGVKKIIRVSGCGAVLPHVKLRDVVVGACTDSKVNRI	120	
Db	65	MGHGVGPSCTSIYTKELITDFGVKKIIRVSGCGAVLPHVKLRDVVVGACTDSKVNRI	124	
QY	121	FKDHFAAIAIDFMRNADVAAGALGIDARVGNLFSADLFYSPGEMFDMVMEKYILGVE	180	
Db	125	FKDHFAAIAIDFMRNADVAAGALGIDARVGNLFSADLFYSPGEMFDMVMEKYILGVE	184	
QY	181	MEAGIYGVAAEFKAKALITCTVSDHINTHROTAAERQTTFNDMIKIALESVLLGDKE	239	
Db	185	MEAGIYGVAAEFKAKALITCTVSDHINTHROTAAERQTTFNDMIKIALESVLLGDKE	243	
RESULT 2				
US-09-543-681A-6543				
; Sequence 6543, Application US/09543681A				
; Patent No. 6605709				
; GENERAL INFORMATION:				
; APPLICANT: GARY BRETON				
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS				
; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS				
; FILE REFERENCE: 2709.1002-001				

```
; CURRENT APPLICATION NUMBER: US/09/543,681A
; CURRENT FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: US 60/128,706
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 6543
; LENGTH: 264
; TYPE: PRT
; ORGANISM: Proteus mirabilis
US-09-543-681A-6543

Query Match      86.3%; Score 1057; DB 4; Length 264;
Best Local Similarity 87.2%; Pred. No. 2e-115;
Matches 205; Conservative 15; Mismatches 15; Indels 0; Gaps 0;

QY 1 MATPHINAENGDFADVVLMFGDPLRAKYIAETFELEDAREVNNVRGMLGFTGTGKGRKISV 60
Db 27 MATPHINAENGDFADVVLMFGDPLRAKYIAETFELEDAREVNNVRGMLGFTGTGKGRKISV 86
QY 61 MGHGVGIPSCSIYTKELITDFGVKKIIRVSGCGAVLPHVKLRDVGWGMGACTDSKNRIR 120
Db 87 MGHGMGIPSCSIYAKELITDFGVKKIIRVSGCGAVLPHVKLRDVGWGMGACTDSKNRLR 146
QY 121 FKQDFAAIAADFQVNRNAVDAKALGIDARVGNLFSADLFYSPDGENFDMVKYGIIGVE 180
Db 147 FKQDFAAIAADFELVQNAVSAKAKDIKVRVGNLFSADLFYSPDPEFDMVKYGIIGVE 206
QY 181 MEAAGIYGVAABFGAKALTCTVSDHIRTHTQTTAABRQTTFNMDIKIALESVLL 235
Db 207 MEAAGIYGVAAYGARALTCTVSDHIKGTQTTSEERQTTFNEMIBIALESVLL 261

RESULT 3
US-09-502-540-10882
; Sequence 10882, Application US/09902540
; Patent No. 6833447
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15849)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; CURRENT FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/217,883
; PRIOR FILING DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 10882
; LENGTH: 221
; TYPE: PRT
; ORGANISM: Myxococcus xanthus
US-09-502-540-10882

Query Match      53.3%; Score 653; DB 4; Length 221;
Best Local Similarity 58.2%; Pred. No. 3.6e-69;
Matches 124; Conservative 39; Mismatches 50; Indels 0; Gaps 0;

QY 19 MPGDPLRAKYIAETFELEDAREVNNVRGMLGFTGTGKGRKISVNGHGVGIPSCSIYTKELI 78
Db 1 MPGDPLRARIYSDRFLLEGAREVTSVRNMLGFTGTGFRGRRVSVNGHGMGVPSPISYATALI 60
QY 79 TDFGVKKIIRVSGCGAVLPHVKLRDVGWGMGACTDSKNRIRFKDHDFAAIAADPDVNRNA 138
Db 61 KTYGVRIIRVSGCGALSTQDKVREVIVATGAGTDSNVNRMLKGHDFAAIVADFTLARRA 120
QY 139 VDAKALGIDARVGNLFSADLFYSPDGENFDMVKYGIIGVEAAGIYGVAABFGAKAL 198
Db 121 MEAARERNKVRAGPVFTSDFYHPQQLNATLARVGLAVENAVAGIYGVAABFGAKAL 180
QY 199 TICTVSDHIHTHQTAAERQTTFNMDIKIALE 231
Db 181 GLLTVSDHIITGESLTPQERQTTFNEMIBIALD 213
```

RESULT 4

```
US-09-107-532A-4839
; Sequence 4839, Application US/09107532A
; Patent No. 6583275
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 7310
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: PC
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,532A
; FILING DATE: 30-Jun-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/085,598
; FILING DATE: 14 May 1998
; APPLICATION NUMBER: 60/051571
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Ariniello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 4839:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 244 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ORIGINAL SOURCE:
; ORGANISM: Enterococcus faecium
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (B) LOCATION 1...244
; SEQUENCE DESCRIPTION: SEQ ID NO: 4839:
US-09-107-532A-4839
```

```
Query Match      49.6%; Score 607; DB 4; Length 244;
Best Local Similarity 52.8%; Pred. No. 1e-62;
Matches 122; Conservative 40; Mismatches 67; Indels 2; Gaps 2;

QY 5 HINAENGDFADVVLMFGDPLRAKYIAETFELEDAREVNNVRGMLGFTGTGKGRKISVNGHG 64
Db 13 HIEAKEGEIADKLLPGDPLRAKYIAETFELEDVPCVYNQVRGMLGFTGKYGERISVQGTG 72
QY 65 VGIPSCSIYTKELITDFGVKKIIRVSGCGAVLPHVKLRDVGWGMGACTDSKNRIRFKDH 124
Db 73 MGNPSATIYAHLEIQSYGVKKLRVGTGALSKDVHVRDLVLAQGAATSSSMIEKQFQAF 132
QY 125 DFAAIAADFQVNRNAVDAKALGIDARVGNLFSADLFYSPD-GEFDMVKYGIIGVEMEA 183
Db 133 HPPPIISDFNLLKAYEIAKEKGYTVHGVNLSSEDSFYKDDLTETFQLAE-LGVLGVEMEA 191
QY 184 AGIYGVAABFGAKALTCTVSDHIRTHTQTTAABRQTTFNMDIKIALESVL 234
Db 192 AALYILGAKYHVQTLMTVSDHLITGBETTAAERQTTFNEMIEVGLETAI 242
```


Db 5 HIAAQGEIADKILLPGDPLRAKFAENFLGDVACFNEVRNMFYGTCTYKGRHVRVMGTG 64
Qy 65 VGPSCSIYTKELITDFGVKKIIRVSCGAVLPHVKLRDVIWGACTDSKVNRIKPDH 124
Db 65 MGMPISIIYARELIVDGVKKIIRVGTAGSLNEBHVHRELVLQAQAATNSIVRNDMPQY 124
Qy 125 DFAAIAIDFDMVRNAVDAAKALGIDARVGNLFSADLFYSPDGMFDMVMEKYGILGVEEAA 184
Db 125 DFQIASFDLLDKAYHIAKELGHTWTHVGNVLSDFVSNYFEKNIELGKKGKAVENEAA 184
Qy 185 GIYGVAAEFAGAKALTICTVSDH-IRTHEQTAAERQTTTFNDMIIKIALESVL 234
Db 185 ALYVLAQYHVDALAIWITSDSLVNPDEDTAERQNTFTDMMKVGLETLI 235

RESULT 8

US-09-134-001C-4239
; Sequence 4239, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; CURRENT APPLICATION NUMBER: GTC-007
; CURRENT FILING DATE: 1998-08-13
; PRIOR FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 4239
; LENGTH: 243
; TYPE: PR1
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-4239

Query Match 46.7%; Score 572; DB 3; Length 243;
Best Local Similarity 53.2%; Pred. No. 1.3e-58;
Matches 123; Conservative 30; Mismatches 76; Indels 2; Gaps 2;
Qy 2 ATPHINAEAGDPAADVLMPPGDLRAKYIAETFLDAREVNNVRGMLGFTGTGKRIKISVM 61
Db 11 STPHINPVGKIAKTVLMPGDLRAKYIAENFLENVQFQTVRNMFGYTGTYKGEVSM 70
Qy 62 GHGVGPSCSIYTKELITDFGVKKIIRVSCGAVLPHVKLRDVIWGACTDSK-VNRI 120
Db 71 GSGMGIPSGIYSYELHYFFDVTIIRVSCGALQEDVNLVDYIIAQAASSTNSNYVDQFN 130
Qy 121 FKDHDFAAIADFDMVRNAVDAAKALGIDARVGNLFSADLFYSPDGMFDMVMEKYGILGVE 180
Db 131 IPGH-FAPIADFNVLVAKAKAADEIGAIHVGNVLSDDTFYNADSTFNDSSKQWGLIGIE 189
Qy 181 MEAAGIYGVAAEFAGAKALTICTVSDHIRTHTHEQTAAERQTTTFNDMIIKIALE 231
Db 190 MESAGLYLNAIHANKALGIFTVSDHILRDEATSAERQTSFTQMMBEIALE 240

RESULT 9

US-09-634-238-214
; Sequence 214, Application US/09634238
; Patent No. 6544772
; GENERAL INFORMATION:
; APPLICANT: Glenn, Matthew
; APPLICANT: Havukala, Ilkka J.
; APPLICANT: Bloksberg, Leonard, N.
; APPLICANT: Lubbers, Mark W.
; APPLICANT: Dekker, James
; APPLICANT: Christenson, Anna C.
; APPLICANT: Holland, Rose
; APPLICANT: O'Toole, Paul W.
; APPLICANT: Reid, Julian R.
; APPLICANT: Coolbear, Timothy

; TITLE OF INVENTION: Polynucleotides, materials incorporating
; TITLE OF INVENTION: them and methods for using them.
; FILE REFERENCE: 11000.1043U1
; CURRENT APPLICATION NUMBER: US/09/634,238
; CURRENT FILING DATE: 2000-08-08
; NUMBER OF SEQ ID NOS: 422
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 214
; LENGTH: 227
; TYPE: PR1
; ORGANISM: Lactobacillus rhamnosus
US-09-634-238-214
Query Match 44.0%; Score 538.5; DB 4; Length 227;
Best Local Similarity 51.1%; Pred. No. 1e-54;
Matches 116; Conservative 39; Mismatches 69; Indels 3; Gaps 3;
Qy 1 MATPHINAEAGDPAADVLMPPGDLRAKYIAETFLDAREVNNVRGMLGFTGTGKRIKISV 60
Db 1 MST-HIDAPKGAIAADVLLPGDPLRAKYIAEHLEKAVRYNTVRNAPGYTGTTEGRRISV 59
Qy 61 MGHGVGPSCSIYTKELITDFGVKKIIRVSCGAVLPHVKLRDVIWGACTDSKVNRI 120
Db 60 QATGMGIPSGIYVNELIQDYGVKTLIRVGTAGMGSDVKVRDVLVQGSSTDSIVLNT 119
Qy 121 F-KDHDFAAIADFDMVRNAVDAAKALGIDARVGNLFSADLFYSPDGMFDMVMEKYGILV 179
Db 120 FGAGMYFAPIAIDFQILUREAANLADAGALRYHGVNGLCEDRFYNDMDRQKLIID-YGVLAT 178
Qy 180 EMEAAGIYGVAAEFAGAKALTICTVSDHIRTHTHEQTAAERQTTTFNDMI 226
Db 179 EMETPALYLLAAKFAQAALSILTNSHLITGSETTAQERQTSFNDMI 225
RESULT 10
US-09-489-039A-8031
; Sequence 8031, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; FILE REFERENCE: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
; CURRENT APPLICATION NUMBER: 2709.2004001
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US/09/489,039A
; PRIOR FILING DATE: 2000-01-27
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 8031
; LENGTH: 298
; TYPE: PR1
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-8031
Query Match 20.0%; Score 245.5; DB 4; Length 298;
Best Local Similarity 28.8%; Pred. No. 3.5e-20;
Matches 72; Conservative 45; Mismatches 102; Indels 31; Gaps 8;
Qy 5 HINAEAGDPAADVLMPPGDLRAKYIAETFLDAREVNNVRGMLGFTGTGKRIK 57
Db 53 HLGLTKNDLQAGATLAIYVPGDPERVEKIAALMDKPKLASHREFTSWRAEL-----DGKP 106
Qy 58 ISVMGHGVGPSCSIYTKELITDFGVKKIIRVSCGAVLPHVKLRDVIWGACTDSKVN 117
Db 107 VIVCGTIGGFTSIAVEEL-AQLGVRTFLRIGTTGAIQIHNVDVLTGTVSGVRLDGS 165
Qy 118 RIRFKDHFAAIADFDMVRNAVDAAKALGIDARVGNLFSADLFYSPDGMFDMV----- 170
Db 166 -LHFAPMEFPAVADFACTTALVEAASIGATTHIGVTASSDTFY-PQERYDTFSGRWS 223
Qy 171 -----MEKY---GILGVMEAGIYGVAAEFAGAKALTICTVSDHIRTHTHEQTAAERQTT 222
Db 224 RFGKSBEEWQAMGVNMYEMESATLLTMCASQGLURAGWAGVIVNRTQQEIPNAETMKQTE 283

```
Qy 223 NDMIKIALES 232
Db 284 SHAVKIVVEA 293

RESULT 11
US-09-543-681A-5709
; Sequence 5709, Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION:
; APPLICANT: GARY BRETON
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
; FILE REFERENCE: 2709.1002-001
; CURRENT APPLICATION NUMBER: US/09/543,681A
; PRIOR FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: US 60/128,706
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 5709
; LENGTH: 265
; TYPE: PRT
; ORGANISM: Proteus mirabilis
US-09-543-681A-5709

Query Match 18.6%; Score 227.5; DB 4; Length 265;
Best Local Similarity 26.4%; Pred. No. 3.8e-18;
Matches 65; Conservative 54; Mismatches 104; Indels 23; Gaps 8;

Qy 5 HINAEKGF--ADVVLMPGDPURAKYIAETPLEDAREVNVVGMGFTGTYGKRKISVNG 62
Db 19 HLGLTKNDLQATLAIVPCDPKRVEKIAK-LMDNPVHLASREYTSWRGEIDGKAVIVCS 77
Qy 63 HGVGTPSCSIYTKELITDFGVKKIIRVSGCGAVLPHVKLRDVIWICMGACTDSKVNRIKPK 122
Db 78 TGIGGPFSTIAVEEL-AQLGRTFURIGTGTGAIQEHINVGDLVLTAAVRLDGAS-LHPA 135
Qy 123 DHDFAAIDFDMVRNAVDAAKALGIDARVGNLFSADLFYSPDGMFV----- 170
Db 136 PWEFPVSDFCFNMALYKAAKONGSTVHVGVTFASDFTY-PQERYDYTTGTVVRRFKGS 194
Qy 171 ---MEKYGILGVEMEAAGIYGVAAEFGAKALTICTVSHIRTHEQTAAE-RQITFNDMI 226
Db 195 MKWQEMGVNVMYEMESATLLTTCASQGLRAGVAGVIYN-RTQQRIPDAELLKKTENNAL 253
Qy 227 KIALES 232
Db 254 GIVIEA 259

RESULT 12
US-09-107-532A-6816
; Sequence 6816, Application US/09107532A
; Patent No. 6583275
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; NUMBER OF SEQUENCES: 7310
; CORRESPONDENCE ADDRESS:
; ADDRESS: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD/ROM ISO9660
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:

Qy 223 NDMIKIALES 232
Db 284 SHAVKIVVEA 293

RESULT 13
US-09-489-039A-12428
; Sequence 12428, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 12428
; LENGTH: 238
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-12428

Query Match 11.1%; Score 136; DB 4; Length 238;
Best Local Similarity 23.0%; Pred. No. 1.7e-07;
Matches 53; Conservative 39; Mismatches 102; Indels 36; Gaps 6;
```

```
; APPLICATION NUMBER: US/09/107,532A
; FILING DATE: 30-Jun-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/085,598
; FILING DATE: 14 May 1998
; APPLICATION NUMBER: 60/051571
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Ariniello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 6816:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 236 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEICAL: YES
; ORIGINAL SOURCE:
; ORGANISM: Enterococcus faecium
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (B) LOCATION 1...236
; SEQUENCE DESCRIPTION: SEQ ID NO: 6816:
US-09-107-532A-6816

Query Match 11.3%; Score 139; DB 4; Length 236;
Best Local Similarity 23.8%; Pred. No. 7.5e-08;
Matches 48; Conservative 37; Mismatches 83; Indels 34; Gaps 5;

Qy 45 GMLGFTGYTKRKISVMGHGVGIPSCSIYTKELITDFGVKKIIRVSGCGAVLPHVKLRDV 104
Db 36 GALFISGLNHEVIVVRSIGIKVLASITSLIIQQYGVNVINTGSAGGIGEGIQVGDI 95
Qy 105 VIGWACTDSKVNRIKPKDHFDAI-----ADFDVNRNAVDAAKALGIDA 149
Db 96 VIS-----DKVAVPDADATGFGVNPGLPGMPLYEASTYLRSEMVAKAATNLNA 146
Qy 150 RVGNLFSADLFYSPDGMFVMEKY-GILGVEMEAAGIYGVAAEFGAKALTICTVSDHIR 208
Db 147 KEGLVITGDTFVDSPKIKELITNPPEALACEMGAAGVQTAQRFNIPFLIVRAMSD--- 203
Qy 209 THEQTAAERQT-TFNDMIKIA 229
Db 204 -----TADHSATQSFDEFIEDA 220

RESULT 13
US-09-489-039A-12428
; Sequence 12428, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 12428
; LENGTH: 238
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-12428

Query Match 11.1%; Score 136; DB 4; Length 238;
Best Local Similarity 23.0%; Pred. No. 1.7e-07;
Matches 53; Conservative 39; Mismatches 102; Indels 36; Gaps 6;
```

```
;
; TYPE: PRT
; ORGANISM: Escherichia coli
; US-09-347-878-18

Query Match      10.6%; Score 130; DB 3; Length 232;
Best Local Similarity 21.3%; Pred. No. 8.3e-07;
Matches 49; Conservative 43; Mismatches 102; Indels 36; Gaps 6;

Qy 32 TFLEDA---REVNNVRGMLGFTGYKGRKISVMGHGVGIPSCSIYTKELITDFGVKKIIR 88
Db 20 TLLRDKIENRITIIYGSEIYTGQLHGVVDVALLKSGIKGVAAAGATLLERCPQDVIIN 79
Qy 89 VSGCGAVLPHVKLRDVIIVGMGACTDSKVNIRIRFKDHFDAAI-----ADFD 133
Db 80 TGSAGGLASTLKVGDIIVS-----DEARYHDADVTAFGYEVGQLPGCPAGFKADEK 130
Qy 134 MVRNADAAKALGIDARVGNLFSADLFYSPDGMFVMEKY-GILGVMEAAAGIYGVAEE 192
Db 131 LVAAAECAIKALDNLNAGRLIVSGDAFINGSVGLAKIRHNPQQAIAVEMEATAIAHVCHN 190
Qy 193 FGAKALTICTVSD-----HIRTHEQTAAERQTTFNDMIKIALESVLLG 236
Db 191 FKVPFVVVRAISDVADQOOSHLSPFEFLAVAAQSTL--MVENLVQNLRG 238

RESULT 14
US-09-543-681A-6066
; Sequence 6066, Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION:
; APPLICANT: GARY BRETON
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.1002-001
; CURRENT APPLICATION NUMBER: US/09/543,681A
; CURRENT FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: US 60/128,706
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 6066
; LENGTH: 242
; TYPE: PRT
; ORGANISM: Proteus mirabilis
; US-09-543-681A-6066

Query Match      10.7%; Score 131; DB 4; Length 242;
Best Local Similarity 22.8%; Pred. No. 6.8e-07;
Matches 47; Conservative 38; Mismatches 85; Indels 36; Gaps 5;

Qy 49 FTGYKGRKISVMGHGVGIPSCSIYTKELITDFGVKKIIRVSGCGAVLPHVKLRDVIIVGM 108
Db 41 YTGKINGVDVALLKSGIKGVAAAGTTLLEHFRPDVIVNTGSGAGLDKALNVGDIIVS- 99
Qy 109 GACTDSKVNIRIRFKDHFDAI-----IADFDMVRNADAAKALGIDARVGN 153
Db 100 -----TEVRVYHDADVTAFGYEPGQWACPPAFIADPKLVNTIAQECIGSLKLNVRGL 151
Qy 154 LFSADLFYSPDGMFVMEKY-GILGVMEAAAGIYGVAEEFGAKALTICTVSD-----H 206
Db 152 ICSGDAFINGAEPLARIRRTPEVVAVEMESTAI GHVCHQPDTPFVVVRAISDVADKESH 211
Qy 207 IRTHEQTAAERQ-----TFENDMIK 227
Db 212 LSFDFELSVAAQQSSSLVTTMLDKLK 237

RESULT 15
US-09-347-878-18
; Sequence 18, Application US/09347878C
; Patent No. 6376210
; GENERAL INFORMATION:
; APPLICANT: Yuan, Chong
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR ASSAYING ANALYTES
; FILE REFERENCE: 25885-1651
; CURRENT APPLICATION NUMBER: US/09/347,878C
; CURRENT FILING DATE: 1999-07-06
; NUMBER OF SEQ ID NOS: 75
; SOFTWARE: PatentIn ver. 2.0
; SEQ ID NO 18
; LENGTH: 232
```

```
;
; TYPE: PRT
; ORGANISM: Escherichia coli
; US-09-347-878-18

Query Match      10.6%; Score 130; DB 3; Length 232;
Best Local Similarity 21.3%; Pred. No. 8.3e-07;
Matches 49; Conservative 43; Mismatches 102; Indels 36; Gaps 6;

Qy 32 TFLEDA---REVNNVRGMLGFTGYKGRKISVMGHGVGIPSCSIYTKELITDFGVKKIIR 88
Db 14 TLLRDKIENRITIIYGSEIYTGQLHGVVDVALLKSGIKGVAAAGATLLLEHCKPQDVIIN 73
Qy 89 VSGCGAVLPHVKLRDVIIVGMGACTDSKVNIRIRFKDHFDAAI-----ADFD 133
Db 74 TGSAGGLAPTILKVGDIIVS-----DEARYHDADVTAFGYEVGQLPGCPAGFKADDK 124
Qy 134 MVRNADAAKALGIDARVGNLFSADLFYSPDGMFVMEKY-GILGVMEAAAGIYGVAEE 192
Db 125 LIAAAEACIAELNLNAGRLIVSGDAFINGSVGLAKIRHNPQQAIAVEMEATAIAHVCHN 184
Qy 193 FGAKALTICTVSD-----HIRTHEQTAAERQTTFNDMIKIALESVLLG 236
Db 185 FNVPFVVVRAISDVADQOOSHLSPFEFLAVAAKQSSL--MVESLVQKLAHG 232

Search completed: July 18, 2005, 21:41:28
Job time : 44 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 18, 2005, 21:40:01 ; Search time 155 Seconds
(without alignments)
598.561 Million cell updates/sec

Title: US-10-035-300A-2

Perfect score: 1225

Sequence: 1 MATPHINAEAGDFADVLMPLP.....TTFNDMIKIALESVLLGDKE 239

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1736639 seqs, 398198149 residues

Total number of hits satisfying chosen parameters: 1736639

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:*

- 1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/2/pubpaa/PCT05_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/2/pubpaa/US10D_PUBCOMB.pep.*
- 17: /cgn2_6/ptodata/2/pubpaa/US10E_PUBCOMB.pep.*
- 18: /cgn2_6/ptodata/2/pubpaa/US10F_PUBCOMB.pep.*
- 19: /cgn2_6/ptodata/2/pubpaa/US11A_PUBCOMB.pep.*
- 20: /cgn2_6/ptodata/2/pubpaa/US11_NEW_PUB.pep.*
- 21: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
- 22: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	DB ID	Description
1	1222	99.8	239	17	US-10-891-096-4
2	693.5	56.6	233	16	US-10-772-272A-16
3	675	55.1	233	9	US-09-895-913A-148
4	663	54.1	231	15	US-10-335-977-8188
5	593	48.4	229	15	US-10-335-977-8187
6	576.5	47.1	237	18	US-10-617-320-4570
7	575.5	47.0	236	17	US-10-472-928-1598
8	572	46.7	243	18	US-10-724-972A-3878
9	538.5	44.0	227	15	US-10-264-213-122
10	456.5	37.3	203	15	US-10-424-599-202583
11	160.5	13.1	262	14	US-10-156-761-14181
					Sequence 4, Appli
					Sequence 16, Appl
					Sequence 148, App
					Sequence 8188, Ap
					Sequence 8187, Ap
					Sequence 4570, Ap
					Sequence 1598, Ap
					Sequence 3878, Ap
					Sequence 122, App
					Sequence 202583,
					Sequence 14181, A

ALIGNMENTS

RESULT 1

US-10-891-096-4
; Sequence 4, Application US/10891096
; Publication No. US20050074857A1
; GENERAL INFORMATION:
; APPLICANT: Araki, Tadashi
; APPLICANT: Miyake, Hitoki
; APPLICANT: Oikawa, Toshihiro
; TITLE OF INVENTION: Method for Producing a Pyrimidine Nucleoside Compound
; FILE REFERENCE: 018765-161
; CURRENT APPLICATION NUMBER: US/10/891,096
; CURRENT FILING DATE: 2004-07-15
; PRIOR APPLICATION NUMBER: JP 2003-199175
; PRIOR FILING DATE: 2003-07-18
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Escherichia coli
US-10-891-096-4

Query Match 99.8%; Score 1222; DB 17; Length 239;
Best Local Similarity 99.6%; Pred. No. 8.9e-127;
Matches 238; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
Qy 1 MATPHINAEAGDFADVLMPLPQDPLRAKYIAETFLDAREVNNVGMGLFTGTGKRKISV 60
Db 1 MATPHINAEAGDFADVLMPLPQDPLRAKYIAETFLDAREVNNVGMGLFTGTGKRKISV 60
Qy 61 MGHGVGIPSCSIYTKELITDFGVKKIIRVSGCGAVLPHVKLRDVVIGNGACTDSKVNR 120
Db 61 MGHGVGIPSCSIYTKELITDFGVKKIIRVSGCGAVLPHVKLRDVVIGNGACTDSKVNR 120

12 146 11.9 233 15 US-10-282-122A-77737 Sequence 77737, A
13 139 11.3 230 15 US-10-282-122A-57936 Sequence 57936, A
14 136 11.1 232 15 US-10-282-122A-60181 Sequence 60181, A
15 135 11.0 233 15 US-10-282-122A-55749 Sequence 55749, A
16 132.5 10.8 232 15 US-10-282-122A-76082 Sequence 76082, A
17 131.5 10.7 232 15 US-10-282-122A-73207 Sequence 73207, A
18 131 10.7 235 15 US-10-282-122A-68956 Sequence 68956, A
19 130 10.6 229 15 US-10-282-122A-58449 Sequence 58449, A
20 130 10.6 232 14 US-10-287-274-397 Sequence 397, App
21 130 10.6 232 15 US-10-282-122A-56421 Sequence 56421, A
22 128.5 10.5 242 18 US-10-994-726-454 Sequence 454, App
23 128.5 10.5 265 18 US-10-282-122A-47199 Sequence 47199, A
24 128.5 10.5 265 18 US-10-994-726-453 Sequence 453, App
25 127.5 10.4 231 15 US-10-282-122A-74420 Sequence 74420, A
26 125 10.2 269 15 US-10-282-122A-76408 Sequence 76408, A
27 124 10.1 231 15 US-10-282-122A-77444 Sequence 77444, A
28 123.5 10.1 231 10 US-09-882-227-458 Sequence 458, App
29 123.5 10.1 231 15 US-10-282-122A-58620 Sequence 58620, A
30 122.5 10.1 231 15 US-10-282-122A-57033 Sequence 57033, A
31 119.5 9.8 231 15 US-10-335-977-9256 Sequence 9256, Ap
32 118.5 9.7 228 15 US-10-282-122A-44334 Sequence 44334, A
33 118.5 9.7 230 9 US-09-815-242-5831 Sequence 5831, Ap
34 118 9.6 228 15 US-10-282-122A-71661 Sequence 71661, A
35 117.5 9.6 233 15 US-10-282-122A-65006 Sequence 65006, A
36 117 9.6 237 15 US-10-282-122A-47110 Sequence 47110, A
37 115 9.4 193 15 US-10-335-977-9255 Sequence 9255, Ap
38 115 9.4 228 15 US-10-282-122A-70780 Sequence 70780, A
39 115 9.4 231 15 US-10-282-122A-72419 Sequence 72419, A
40 114.5 9.3 233 15 US-10-282-122A-65840 Sequence 65840, A
41 114.5 9.3 290 15 US-10-282-122A-44569 Sequence 44569, A
42 113.5 9.3 262 15 US-10-282-122A-50255 Sequence 50255, A
43 110.5 9.0 231 15 US-10-282-122A-46391 Sequence 46391, A
44 110 9.0 230 15 US-10-282-122A-73938 Sequence 73938, A
45 110 9.0 230 17 US-10-472-928-1958 Sequence 1958, Ap

Qy 121 FKQDFAAIAIDFDMVRNAVDAAKALGIDARVGNLFSADLFYSPDGMFDMVEKYGLGVE 180
Db 121 FKQDFAAIAIDFDMVRNAVDAAKALGIDARVGNLFSADLFYSPDGMFDMVEKYGLGVE 180
Qy 181 MEAGIYGVAEFGAKALTICTVSDHIRTHTQTTAAERQTTFNMDIKIALESVLLGDK 239
Db 181 MEAGIYGVAEFGAKALTICTVSDHIRTHTQTTAAERQTTFNMDIKIALESVLLGDK 239

RESULT 2

US-10-772-272A-16
; Sequence 16, Application US/10772272A
; Publication No. US20040166575A1
; GENERAL INFORMATION:
; APPLICANT: Ajinomoto Co., Inc.
; TITLE OF INVENTION: Inosine producing bacterium belonging to the genus
; TITLE OF INVENTION: Bacillus and method for producing inosine
; FILE REFERENCE: US-108
; CURRENT APPLICATION NUMBER: US/10/772,272A
; CURRENT FILING DATE: 2002-02-06
; PRIOR APPLICATION NUMBER: JP 2003-37760
; PRIOR FILING DATE: 2003-02-17
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 16
; LENGTH: 233
; TYPE: PRT
; ORGANISM: Bacillus subtilis
US-10-772-272A-16

Query Match 56.6%; Score 693.5; DB 16; Length 233;
Best Local Similarity 58.1%; Pred. No. 3.6e-68;
Matches 133; Conservative 41; Mismatches 54; Indels 1; Gaps 1;

Qy 5 HINAEAGDFAVLMPCGDLRAKYAETFLDAREVNNVRGMLGFTGTYGKRKISVNGHG 64
Db 4 HIGAEKGIAIDVLLPGDPLRAKYAETYLENVCEYNEVRGMYGFTGTYGKRKISVQGTG 63
Qy 65 VGIPSCSIYTKELITDFGVKKIIRVSGCAVLPHVKLRDVGVCACDTSKVNRIKDH 124
Db 64 MGVPESIIYVNELIQSDVQNLIRVSGCAIRKDKVRDVLAMTSDTSQMNRAFGSV 123
Qy 125 DFAAIAIDFDMVRNAVDAAKALGIDARVGNLFSADLFYSPDGMFDMVEKYGLGVEEAA 184
Db 124 DFACADFELKKNAYDAKOKGVPTVGSVFTAQDFNDSQI-EKLAQVGLGVEMETT 182
Qy 185 GIYGVAAEFGAKALTICTVSDHIRTHTQTTAAERQTTFNMDIKIALESV 233
Db 183 ALYTLAAKHGRKALSILTVSDHVLGTBETTAERQTTTFHDMIEVALHSV 231

RESULT 3

US-09-895-913A-148
; Sequence 148, Application US/09895913A
; Patent No. US20020160456A1
; GENERAL INFORMATION:
; APPLICANT: Kleanthous, Harold
; APPLICANT: Al-Garawi, Amal
; APPLICANT: Miller, Charles
; APPLICANT: Tomb, Jean Francois
; APPLICANT: Oomen, Raymond P.
; TITLE OF INVENTION: Identification of Polynucleotides
; TITLE OF INVENTION: Encoding No. US20020160456A1 Helicobacter Polypeptides in the
; TITLE OF INVENTION: Genome
; FILE REFERENCE: 06132/043002
; CURRENT APPLICATION NUMBER: US/09/895,913A
; CURRENT FILING DATE: 2001-06-29
; PRIOR APPLICATION NUMBER: US 08/881,227
; PRIOR FILING DATE: 1997-06-24
; NUMBER OF SEQ ID NOS: 368
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 148

; LENGTH: 233
; TYPE: PRT
; ORGANISM: Helicobacter pylori
US-09-895-913A-148

Query Match 55.1%; Score 675; DB 9; Length 233;
Best Local Similarity 55.0%; Pred. No. 4.1e-66;
Matches 126; Conservative 42; Mismatches 61; Indels 0; Gaps 0;

Qy 3 TPHINAEAGDFAVLMPCGDLRAKYAETFLDAREVNNVRGMLGFTGTYGKRKISVNG 62
Db 2 TPHINAKIGDFYPCQLLGGDLRVSIAKFLQDAKEITNVRNMLGFSYKGRGSLMG 61
Qy 63 HGVGIPSCSIYTKELITDFGVKKIIRVSGCAVLPHVKLRDVGVCACDTSKVNRIKDH 122
Db 62 HGVGIASCTIYVTELLKTYQVKELLAIGTCGALSPKVLKDIIMATGASTDSKTNVRFL 121
Qy 123 DHPFAAIAIDFDMVRNAVDAAKALGIDARVGNLFSADLFYSPDGMFDMVEKYGLGVE 182
Db 122 NHDLSATPDPELSIRAYQTAKRLGIDLKVGNVFSSDPFYSFETHAFDLMAKYNHLAEME 181
Qy 183 AAGIYGVAEFGAKALTICTVSDHIRTHTQTTAAERQTTFNMDIKIALE 231
Db 182 AAGLYATAMELNAKALCLCSVDHLITKEALSPKRVESPDNMIILALE 230

RESULT 4

US-10-335-977-8188
; Sequence 8188, Application US/10335977
; Publication No. US20040052799A1
; GENERAL INFORMATION:
; APPLICANT: DOUGLAS SMITH et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
; RELATING TO HELICOBACTER PYLORI FOR
; DIAGNOSTICS AND THERAPEUTICS

; NUMBER OF SEQUENCES: 10031
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 28 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-1875
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: Windows NT 4.0
; SOFTWARE: UNIX
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/335,977
; FILING DATE: 30-Dec-2002
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/993,002
; FILING DATE: 17-DEC-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Mandragouras, Amy E.
; REGISTRATION NUMBER: 36,207
; REFERENCE/DOCKET NUMBER: CTN-018
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)227-7400
; TELEFAX: (617)742-4214
; INFORMATION FOR SEQ ID NO: 8188:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 231 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEetical: YES
; ORIGINAL SOURCE:
; ORGANISM: Helicobacter pylori
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1...231

SEQUENCE DESCRIPTION: SEQ ID NO: 8188;
US-10-335-977-8188

Query Match 54.1%; Score 663; DB 15; Length 231;
Best Local Similarity 53.7%; Pred. No. 8,6e-65;
Matches 123; Conservative 44; Mismatches 62; Indels 0; Gaps 0;
Qy 3 TPHNAEAGDPAVVLMPGDFPLRAKYIAETFLDAREVNNVRGMLGFTGYKGRKISVMG 62
Db 2 TPHNAKIGDFYPCCLCGDFPLRVSYIAKNFLQDAKEITNVRNMLGFGSKYKGGISLMG 61
Qy 63 HGVIPSCSIYTKELITDFGVKKIIRVSCCAVLPVHVKLRDVIOMGACTDSKVNRFK 122
Db 62 HGMGIASCTIYVTELKTYQKELLRIGTCGAISPKVGLKDIYMATGASTDSKTRVFL 121
Qy 123 DHDFAAADDFVNRNAVDAKALGIDARVGNLFSADLFSYSPDGEFVMEKYGILGVEME 182
Db 122 NHDLSATPDFLSLRVQTKRGLIDLKIGNVFSDFYSFETHAFLMAQYNHLAIME 181
Qy 183 AAGYGVAAEFGAKALTICTVSDHRTHEQTAAERQTFNMDIKIALE 231
Db 182 AAGLYATAMELNAKALCLCSVDHLITKEALSPKRIESFDNMITLALE 230

RESULT 5

US-10-335-977-8187
; Sequence 8187, Application US/10335977
; Publication No. US20040052799A1
; GENERAL INFORMATION:

APPLICANT: DOUGLAS SMITH et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
RELATING TO HELICOBACTER PYLORI FOR
DIAGNOSTICS AND THERAPEUTICS

NUMBER OF SEQUENCES: 10031
CORRESPONDENCE ADDRESS:
ADDRESSEE: LAHIVE & COCKFIELD
STREET: 28 State Street
CITY: Boston
STATE: Massachusetts
COUNTRY: USA
ZIP: 02109-1875
COMPUTER READABLE FORM:
MEDIUM TYPE: CD/ROM ISO9660
COMPUTER: IBM PC Compatible
OPERATING SYSTEM: Windows NT 4.0
SOFTWARE: UNIX
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/335,977
FILING DATE: 30-Dec-2002

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/993,002
FILING DATE: 17-DEC-1997
ATTORNEY/AGENT INFORMATION:
NAME: Mandragouras, Amy E.
REGISTRATION NUMBER: 36,207
REFERENCE/DOCKET NUMBER: GTN-018
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)227-7400
TELEFAX: (617)742-4214

INFORMATION FOR SEQ ID NO: 8187:
SEQUENCE CHARACTERISTICS:

LENGTH: 229 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

HYPOTHETICAL: YES

ORIGINAL SOURCE:

ORGANISM: Helicobacter pylori

FEATURE:

NAME/KEY: misc feature

LOCATION: (B) LOCATION 1...229

SEQUENCE DESCRIPTION: SEQ ID NO: 8187:

US-10-335-977-8187

Query Match 48.4%; Score 593; DB 15; Length 229;
Best Local Similarity 52.4%; Pred. No. 5e-57;
Matches 110; Conservative 40; Mismatches 60; Indels 0; Gaps 0;
Qy 22 DPLRAKYIAETFLDAREVNNVRGMLGFTGYKGRKISVMGHGVIPSCSIYTKELITDF 81
Db 20 DPFVSYIARYLLQDAKEITNVRNMLGFGSKYKGGISLMGHGMIASCTIYVTELKTY 79
Qy 82 GVKIIRVSGCAVLPVHVKLRDVIIVGACATDSKVNRIKPDHFAAIAADFDVNRNAVDA 141
Db 80 QVKELLRIGTCGALSPKVGLKDIYMATGASTDSKTRVRFNLHDLSATPDLSLRAYQT 139
Qy 142 AKALGIDARVGNLFSADLFSYSPDGEFVMEKYGILGVEMEAAAGIYGVAAEFGAKALTIC 201
Db 140 AKRLGIDLKIGNVFSDFYSFETHAFLMAQYNHLAIMEAAGLYATAMELNAKALCLC 199
Qy 202 TVSDHRTHEQTAAERQTFNMDIKIALE 231
Db 200 VSDHLITKEALSPKRIESFDNMITLALE 229

RESULT 6

US-10-617-320-4570
; Sequence 4570, Application US/10617320
; Publication No. US20050136404A1
; GENERAL INFORMATION:

APPLICANT: Lynn A Doucette-Stamm and David Bush
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID
SEQUENCES RELATING TO STREPTOCOCCUS PNEUMONIAE
THERAPEUTICS

NUMBER OF SEQUENCES: 5206

CORRESPONDENCE ADDRESS:

ADDRESSEE: GENOME THERAPEUTICS CORPORATION

STREET: 100 Beaver Street

CITY: Waltham

STATE: Massachusetts

COUNTRY: USA

ZIP: 02354

COMPUTER READABLE FORM:

MEDIUM TYPE: CD/ROM ISO9660

COMPUTER: <Unknown>

OPERATING SYSTEM: <Unknown>

SOFTWARE: <Unknown>

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/617,320

FILING DATE: 10-Jul-2003

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/09/107,433

FILING DATE: 30-Jun-1998

APPLICATION NUMBER: 60/ 085131

FILING DATE: May 12, 1998

APPLICATION NUMBER: 60/051553

FILING DATE: July 2, 1997

ATTORNEY/AGENT INFORMATION:

NAME: Ariniello, Pamela Deneke

REGISTRATION NUMBER: 40,489

REFERENCE/DOCKET NUMBER: GTC-011

TELECOMMUNICATION INFORMATION:

TELEPHONE: (781)893-5007

TELEFAX: (781)893-8277

INFORMATION FOR SEQ ID NO: 4570:

SEQUENCE CHARACTERISTICS:

LENGTH: 237 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

HYPOTHETICAL: YES

ORIGINAL SOURCE:

ORGANISM: Streptococcus pneumoniae

FEATURE:

NAME/KEY: misc feature

LOCATION: (B) LOCATION 1...237

;

FOR DIAGNOS


```
Db 1 MST-HIDAPKGAIDVLLPGDPLRAQVIAEHFLEKAVRYNTRNNAFCYTGTFEGRRISV 59
Qy 61 MGHGVGPSCSIYTKELITDFGVKKIIRVSGCAGVLPVHKLRDVVVGMACTDSKVNRI 120
Db 60 QATGWGIPSIYVNELIYGVKTLIRVGTAGGSDVKVRDVLVQSSSTDSSVLNT 119
Qy 121 F-KHDFPAAIADPDMVRNADAAKALGIDARVGNLFSADLFSYSPDGMFDMVEKYGILCV 179
Db 120 FGAGMYFAPADFDQLLRANLADAGALRYHVGNVGLGDFYNDMDRQKLID-YGVLAT 178
Qy 180 EMEAGIYGVAAEFGAKALITCTVSDHIRTHTOQTAAERQTTFNMDI 226
Db 179 EMEPALVLLAAKFAQAALSILTVSNHLITGEETTAQERQTSFNMDI 225

RESULT 10
US-10-424-599-202583
; Sequence 202583, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 202583
; LENGTH: 203
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_24958C.1.pep
US-10-424-599-202583

Query Match 37.3%; Score 456.5; DB 15; Length 203;
Best Local Similarity 45.9%; Pred. No. 5.8e-42;
Matches 106; Conservative 24; Mismatches 66; Indels 35; Gaps 3;

Qy 2 ATPHINAEAGDPADVLMGPDPLRAKYIAETFLDAREVNVRGMLGFTGYTKGRKISVM 61
Db 4 STPHINPENGVIKATVLMGPDPLRAQYIAENFLENVEQFNTVRNMFYGTGYTKGEVSM 63
Qy 62 GHGVGPSCSIYTKELITDFGVKKIIRVSGCAGVLPVHKLRDVVVGMACTDSK-VNRIR 120
Db 64 GSGMGIPSIGIYSELVYHFFDVTIIRVSGCALQEDVNLVDVIAQAASNTSNVVDQFN 123
Qy 121 FKDHDFAAIADPDMVRNADAAKALGIDARVGNLFSADLFSYSPDGMFDMVEKYGILGVE 180
Db 124 IPGH-FAPADFNLVAKAKKADEIGALSHV----- 153

Qy 181 MEAGIYGVAAEFGAKALITCTVSDHIRTHTOQTAAERQTTFNMDIKIALE 231
Db 154 ---GLVNLNATHANKALGIFTVSDHILRDEATSAERQTSFTQWMEIALE 200

RESULT 11
US-10-156-761-14181
; Sequence 14181, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
```

```
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 14181
; LENGTH: 262
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-14181

Query Match 13.1%; Score 160.5; DB 14; Length 262;
Best Local Similarity 30.3%; Pred. No. 5.7e-09;
Matches 56; Conservative 31; Mismatches 91; Indels 7; Gaps 5;

Qy 13 PADVLMGPDPLRAKYIAETFLDAREVNVRGMLGFTGYTKGRKISVMGHGVGPSCSI 72
Db 19 YAVVV---GDPARAAAVA-ALLDGAEVSHYREYRVFRGAWKGLPVTVASHGVGPGAIL 74
Qy 73 YTKELITDFGVKKIIRVSGCAGVLPVHKLRDVVVGMACTDSKVNRIKDHDFAAIADF 132
Db 75 LFQEL-ADAGVTFVFRGTAGAKPGIGDGLVIAEAAVRRDGVTO-QLLPAAYPAVSAP 132
Qy 133 DMVRNADVAAKALGIDARVGNLFSADLFSYSPDGMFDMVEKYGILGVMEAGIYGVAAE 192
Db 133 EAVLALQRAARETGAPHHRGIVVWRAAP-QPGLPLDGYARAGLAAIEMELLSALLVTSAL 191
Qy 193 FGAKA 197
Db 192 RGLVA 196

RESULT 12
US-10-282-122A-77737
; Sequence 77737, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zykkind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
```

; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 77737
; LENGTH: 233
; TYPE: PRT
; ORGANISM: Yersinia pestis
US-10-282-122A-77737

Query Match 11.9%; Score 146; DB 15; Length 233;
Best Local Similarity 23.8%; Pred. No. 2e-07;
Matches 54; Conservative 39; Mismatches 98; Indels 36; Gaps 6;
QY 32 TFLDAREVNNRGMGLG---FTGYKGRKISVMGHGVGIPSCSIYTKELITDFGVKKIIR 88
DB 14 TLLRDRIENRQTLARACCEIYTGQNGIDVALLKSGIKGVAAAMGTTLLLEHCQPDIVN 73
QY 89 VSGGAVLPHVKLRDVGIGGACTDSKVNIRIRPKDHPAA-----IADED 133
DB 74 TGSAGGLDSSLLKGVDIVVS-----NEVRHDADVTAAGVEPGQVAGCPAAFVADED 124
QY 134 MVRNAVDAAKALGIDARVGNLFSADLFYSPDGEFDMVEKY-GILGVMEAAAGIYGVAAE 192
DB 125 LIALAENCIQQLKLNARGLICSGDAFINGAEPLARIRAAFPPTVAAVEMEAAIGHVCYL 184
QY 193 FGAKALITCTVSD-----HIRTQTTAAERQTTFNDMIKIALESV 233
DB 185 FNTPFVVVRAISDVADQASHLSFEFLVVAAKQSTL--MIKAMLTTL 229

RESULT 13

US-10-282-122A-57936
; Sequence 57936, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.

; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA 034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614

; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 57936
; LENGTH: 230
; TYPE: PRT
; ORGANISM: Enterococcus faecium
US-10-282-122A-57936

Query Match 11.3%; Score 139; DB 15; Length 230;
Best Local Similarity 23.8%; Pred. No. 1.1e-06;
Matches 48; Conservative 37; Mismatches 83; Indels 34; Gaps 5;
QY 45 GMLGFTGYKGRKISVMGHGVGIPSCSIYTKELITDFGVKKIIRVSGGAVLPHVKLRDV 104
DB 30 GALTFIGSLGNHBEIVVRSGIGKVLASITTSLLIQYGVNMVINTSGAGIGSLQVGDI 89
QY 105 VIGMGACTDSKVNIRIRPKDHPAAI-----ADFDVMVRNAVDAAKALGIDA 149
DB 90 VIS-----DKVAVFDADATGFGYNPQQLPGMPLYYEASTYLRSEMVKAATNLNA 140
QY 150 RVGNLFSADLFYSPDGEFDMVEKY-GILGVMEAAAGIYGVAAEFGAKALITCTVSDHIR 208
DB 141 KEGLIVTGDFTVDSPPKIKEILTNPPEALACEMEAGAAVGTARQFNIPFLIVRAMSD--- 197
QY 209 THEQTTAAERQT-TFNDMIKIA 229
DB 198 -----TADHSATQSFDEFIEDA 214

RESULT 14

US-10-282-122A-60181
; Sequence 60181, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA 034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614

```
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 60181
; LENGTH: 232
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae
US-10-282-122A-60181

Query Match      11.1%; Score 136; DB 15; Length 232;
Best Local Similarity 23.0%; Pred. No. 2.5e-06;
Matches 53; Conservative 39; Mismatches 102; Indels 36; Gaps 6;

Qy 32 TFLDGA---REVNNVRGMLGFTGYKGRKISVMGHGVGIPSCSIYTKELIITDFGVKKIIR 88
Db 14 TLLRDKIENRQTITIGSGEITYTQGLHGVDAVALLKSGIGKVAAMGATLLERCCPDVVIIN 73
Qy 89 VSGCGAVLPHVKLRDVIIVGMACTDSKVNIRIRFKDHPAAI-----ADPFD 133
Db 74 TGSAGGLASTLKVGDIVVS-----DEARYHDADVTAFGYEGQLPGCPAGFAKDEK 124
Qy 134 MYRNAVDAAKALGIDARVGNLFSADLFVSPDGEFDMVEKY-GILGVEMEAAGIYGVAAE 192
Db 125 LVAAAEKICLDLNAVRGLIVSGDAFINGSVGLAKIRHNPQAIIVEMEAATAIAHVCHN 184
Qy 193 FGAKALTICTVSD-----HIRTHEQTAAERQTTFNDMIKIALESVLLG 236
Db 185 FKVPFVVVRAISDVADQOQSHLSPEEFLAARQSTL--MVENLVQNLRG 232

; LENGTH: 233
; TYPE: PRT
; ORGANISM: Enterobacter cloacae
US-10-282-122A-55749

Query Match      11.0%; Score 135; DB 15; Length 233;
Best Local Similarity 22.2%; Pred. No. 3.2e-06;
Matches 51; Conservative 42; Mismatches 101; Indels 36; Gaps 6;

Qy 32 TFLDGA---REVNNVRGMLGFTGYKGRKISVMGHGVGIPSCSIYTKELIITDFGVKKIIR 88
Db 15 TLLRDKIENRQTITIGSGEITYTQGLHGVDAVALLKSGIGKVAALGATLLERCCPDVIVN 74
Qy 89 VSGCGAVLPHVKLRDVIIVGMACTDSKVNIRIRFKDHPAAI-----ADPFD 133
Db 75 TGSAGGLAPTLLKVGDIIVVS-----DEARYHDADVTAFGYEGQLPGCPAGFAKDDK 125
Qy 134 MYRNAVDAAKALGIDARVGNLFSADLFVSPDGEFDMVEKY-GILGVEMEAAGIYGVAAE 192
Db 126 LIAAAETCTIAELNLNAVRGLIVSGDAFINGSVGLAKIRHNPQAVAVEMEAATAIAHVCHN 185
Qy 193 FGAKALTICTVSD-----HIRTHEQTAAERQTTFNDMIKIALESVLLG 236
Db 186 FSVPFVVVRAISDVADQOQSHLSFDEFLTVAARQSTV--MVERLVQNLRG 233

Search completed: July 18, 2005, 21:53:21
Job time : 157 secs

RESULT 15
US-10-282-122A-55749
; Sequence 55749, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA 034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 55749
```

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: July 22, 2005, 22:31:58 ; Search time 5335 Seconds
(without alignments)
2170.721 Million cell updates/sec

Title: US-10-035-300A-2

Perfect score: 1225

Sequence: 1 MATPHINAEMGDFADVVLMP.....TTFNDMIKIALESVLLGDKE 239

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO.spool_p/US10035300/runat_18072005_165350_27973/app_query.fasta_1.391
-DB=GenEmbl -QFMT=fastcap -SUFFIX=rge -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=spect -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10035300@cgn 1.1 5600 @runat_18072005_165350_27973 -NCPU=6 -ICPU=3
-NO MMAP -LARGQUERY -NEG SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FCAPOP=6
-FCAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl.*
1: gb_ba.*
2: gb_htg.*
3: gb_in.*
4: gb_on.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pl.*
9: gb_pt.*
10: gb_ro.*
11: gb_ats.*
12: gb_sy.*
13: gb_un.*
14: gb_vi.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1222	99.8	720	6	BD187724 A method
2	1222	99.8	720	6	CQ774461 Sequence
3	1222	99.8	720	6	AX087917 Sequence
4	1222	99.8	720	6	AX590437 Sequence

ALIGNMENTS

RESULT	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT	
BD187724	BD187724	A method of producing a cytosine nucleoside compound.	BD187724	BD187724.1	GI:32997463	JP 2003018997-A/3.	Escherichia coli	Escherichia coli	Escherichia coli	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Escherichia.	1 (bases 1 to 720)		
										AKAKI, T., IKEDA, I., MATOISHI, K., ABE, R., OIKAWA, T., MATSUBA, Y., NAGAHARA, K., FUKUIRI, Y. and ISHIBASHI, H.	A method of producing a cytosine nucleoside compound	MITSUI CHEMICALS INC	OS Escherichia coli
												PN JP 2003018997-A/3	
												PD 21-JAN-2003	
												PF 01-MAY-2002 JP 2002129867	
												PI TADASHI ARAKI, ICHIRO IKEDA, KAORI MATOISHI, REIKO ABE, TOSHIHIRO OIKAWA,	
												PI YASUKO MATSUBA, KIYOTERU NAGAHARA, YASUSHI FUKUIRI, HIROKI ISHIBASHI	
												PC C12N15/09, C12N1/15, C12N1/19, C12N1/21, C12N5/10, C12N9/10 PC	
												PC C12P19/40, C12N15/00, C12N5/00	

CC A method of producing a cytosine nucleoside compound FH Key
FT source Location/Qualifiers
FT source 1..720 /organism="Escherichia coli".
FEATURES
source Location/Qualifiers
1..720 /organism="Escherichia coli"
/mol_type="genomic DNA"
/db_xref="taxon:562"
ORIGIN
Alignment Scores: 4.06e-110 Length: 720
Pred. No.: 1222.00 Matches: 238
Score: 100.00% Conservativeness: 1
Percent Similarity: 100.00% Mismatches: 0
Best Local Similarity: 99.58% Indels: 0
Query Match: 99.76% Gaps: 0
DB: 6
US-10-035-300A-2 (1-239) x BD187724 (1-720)
Qy 1 MetAlaThrProHisIleAsnAlaGluMetGlyAspPheAlaAspValValLeuMetPro 20
Db 1 ATGGCTACCCACACATTAATGCAGAAATGGCGATTTCGCTGACGTAGTTTTCATGCCA 60
Qy 21 GlyAspProLeuArgAlaIleValThrIleAlaGluThrPheLeuGluAspAlaArgGluVal 40
Db 61 GGCGACCCGCTGGTGGAGATATATTCGTAACCTTTCCTTGAAGATGCCGTAAGTG 120
Qy 41 AsnAsnValArgGlyMetLeuGlyPheThrGlyThrIleValGlySerGlyValAlaValLeuProHisVal 60
Db 121 AACACGTTCCGCGTATGCTGGGCTTACCCGTTACTTACAAAGCGCCGCAAAATTCGCTA 180
Qy 61 MetGlyHisGlyValGlyIleProSerCysSerIleThrIleValGlyLeuLeuThrAsp 80
Db 181 ATGGGTACCGGTATGGGTATCCGCTGCTCATCTACCAAGAACTGATCACCAGAT 240
Qy 81 PheGlyValIleValIleAlaArgValGlySerCysGlyValAlaValLeuProHisVal 100
Db 241 TTCGGCGTGAAGAAATATATCCGGTGGGTTCCTGTGGCGAGTTCTGCCGCGACGTAAA 300
Qy 101 LeuArgAspValValIleGlyMetGlyAlaCysThrAspSerIleValAsnArgIleArg 120
Db 301 CTGCGCGACGTCGTTATCGGTATGGGTGCTGCACCGATTCCAAAAGTTAACCGCATCGAT 360
Qy 121 PheLysAspHisAspPheAlaIleAlaAspPheAspMetValArgAsnAlaValAsp 140
Db 361 TTTAAAGACCATGACTTTGCCGCTATCGCTGACTTCGACATGGTGGCGTAACCGCATAGAT 420
Qy 141 AlaAlaLysAlaLeuGlyIleAspAlaArgValGlyAsnLeuPheSerAlaAspLeuPhe 160
Db 421 GCAGCTAAAGACACTGGGTATGTATGCTCGCGTGGGTAACTGTCTCCGCTGACCTGTTTC 480
Qy 161 TyrSerProAspGlyGluMetPheAspValMetGluLysThrGlyIleLeuGlyValGlu 180
Db 481 TACTCTCCGCGACGCGAAATGTTCCGAGTGCATGAGAAATACGCAATTCCTCGCGCGTGA 540
Qy 181 MetGluAlaAlaGlyIleThrGlyValAlaAlaGluPheGlyAlaLysAlaLeuThrIle 200
Db 541 ATGGAACGCGCTGGTATCTACGCGCTGCTGCAGAAATTTGGCGCGAAAGCCCTGACCATC 600
Qy 201 CysThrValSerAspHisIleArgThrHisGluGlnThrThrAlaAlaGluArgGlnThr 220
Db 601 TGCACCGTATCTGACCAACATCCGACATCAAGACGACACCACTGCCGCTGAGCGTCAGACT 660
Qy 221 ThrPheAsnAspMetIleLysIleAlaLeuGluSerValLeuLeuGlyAspLysGlu 239
Db 661 ACCTTCAACGACATGATCAAAATCGCACTGGAATCCGTTCTGCTGGCGGATAAAGAG 717
RESULT 2
CQ774461 720 bp DNA linear PAT 06-MAR-2004
LOCUS CQ774461
DEFINITION Sequence 23 from Patent WO2004013333.

CQ774461
CQ774461.1 GI:45237695
Escherichia coli
Escherichia coli
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Escherichia.
1
Kock, M., Frank, M. and Badur, R.
Novel selection method
Patent: WO 200401333-A 23 12-FEB-2004;
EASF Plant Science GmbH (DE)
FEATURES
source Location/Qualifiers
1..720 /organism="Escherichia coli"
/mol_type="unassigned DNA"
/db_xref="taxon:562"
1..720 /notes="unnamed protein product; coding for purine
nucleoside phosphorylase (deod)"
/codon_start=1
/transl_table=11
/protein_id="CAF34364.1"
/db_xref="GI:45237695"
/translation="MATPHINAEMGDADVLMFGDPLRAKYIAETFLDAREVNNVR
GMLVGTGKGRKISVMHGMGTPSCSIYTKELITDFGVKKIIRVSGCGAVLPHVKLR
DIVVNGACTDSKVNEIRFKDHPFAIADPDVNRNADAKALGIDARVGNLFSADLF
YSPDGMFDMKYGILGVMEAAAGIYVNAEFGAKALITCTVSDHIRTHEQTAAER
QTTFNMDIKIALESVLLGDKE"
ORIGIN
Alignment Scores: 4.06e-110 Length: 720
Pred. No.: 1222.00 Matches: 238
Score: 100.00% Conservativeness: 1
Percent Similarity: 100.00% Mismatches: 0
Best Local Similarity: 99.58% Indels: 0
Query Match: 99.76% Gaps: 0
DB: 6
US-10-035-300A-2 (1-239) x CQ774461 (1-720)
Qy 1 MetAlaThrProHisIleAsnAlaGluMetGlyAspPheAlaAspValValLeuMetPro 20
Db 1 ATGGCTACCCACACATTAATGCAGAAATGGCGATTTCGCTGACGTAGTTTTCATGCCA 60
Qy 21 GlyAspProLeuArgAlaIleValThrIleAlaGluThrPheLeuGluAspAlaArgGluVal 40
Db 61 GGCGACCCGCTGGTGGAGATATATTCGTAACCTTTCCTTGAAGATGCCGTAAGTG 120
Qy 41 AsnAsnValArgGlyMetLeuGlyPheThrGlyThrIleValGlySerGlyValIleSerVal 60
Db 121 AACACGTTCCGCGTATGCTGGGCTTACCCGTTACTTACAAAGCGCCGCAAAATTCGCTA 180
Qy 61 MetGlyHisGlyValGlyIleProSerCysSerIleThrIleValGlyLeuLeuThrAsp 80
Db 181 ATGGGTACCGGTATGGGTATCCGCTGCTCATCTACCAAGAACTGATCACCAGAT 240
Qy 81 PheGlyValIleValIleAlaArgValGlySerCysGlyValAlaValLeuProHisVal 100
Db 241 TTCGGCGTGAAGAAATATATCCGGTGGGTTCCTGTGGCGAGTTCTGCCGCGACGTAAA 300
Qy 101 LeuArgAspValValIleGlyMetGlyAlaCysThrAspSerIleValAsnArgIleArg 120
Db 301 CTGCGCGACGTCGTTATCGGTATGGGTGCTGCACCGATTCCAAAAGTTAACCGCATCGT 360
Qy 121 PheLysAspHisAspPheAlaIleAlaAspPheAspMetValArgAsnAlaValAsp 140
Db 361 TTTAAAGACCATGACTTTGCCGCTATCGCTGACTTCGACATGGTGGCGTAACCGCATAGAT 420
Qy 141 AlaAlaLysAlaLeuGlyIleAspAlaArgValGlyAsnLeuPheSerAlaAspLeuPhe 160
Db 421 GCAGCTAAAGACACTGGGTATGTATGCTCGCGTGGGTAACTGTCTCCGCTGACCTGTTTC 480

```
Qy 161 TyrSerProAspGlyGluMetPheAspValMetGluLysTyrGlyIleLeuGlyValGlu 180
Db 481 TACTTCGCGAGCGGCAATATGTCACGTGATGGAAAAATACGGCATTCCTCGCGGTGAA 540
Qy 181 MetGluAlaAlaGlyIleTyrGlyValAlaAlaGluPheGlyAlaLysAlaLeuThrIle 200
Db 541 ATGGAGCGGCTGGTATCTACGGCTCGCTGAGAAATTTGGCGCAAGCCCTGACCATC 600
Qy 201 CysThrValSerAspHisIleArgThrHisGluGlnThrThrAlaAlaGluArgGlnThr 220
Db 601 TGCACCGTATCTGACCATCGCACTCAGGAGCAGACCACTCGCGCTGAGCGTCAGACT 660
Qy 221 ThrPheAsnAspMetIleLysIleAlaLeuGluSerValLeuLeuGlyAspLysGlu 239
Db 661 ACCTTCAACGACATGATCAAAATCGCACTGGAATCCGTTCTGCTGGCGGATAAAGAG 717

RESULT 3
AX087917 LOCUS AX087917 720 bp DNA linear PAT 17-MAR-2001
DEFINITION Sequence 3 from Patent WO0114566.
ACCESSION AX087917
VERSION AX087917.1 GI:13396895
KEYWORDS
SOURCE Escherichia coli
ORGANISM Escherichia coli
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Escherichia.
REFERENCE 1
AUTHORS Tischer, W., Ihlenfeldt, H.G., Barzu, O., Sakamoto, H., Pistotnik, E.,
Marliere, P., and Pochet, S.
TITLE Enzymatic synthesis of deoxyribonucleosides
JOURNAL Patent: WO 0114566-A 3 01-MAR-2001;
Roche Diagnostics GmbH (DE); INSTITUT PASTEUR (FR); Pharma-
Waldhof GmbH & Co. KG (DE)
FEATURES
source Location/Qualifiers
1..720
/organism="Escherichia coli"
/mol_type="unassigned DNA"
/db_xref="taxon:562"
1..720
/feature="unnamed protein product"
/codon_start=1
/transl_table=11
/protein_id="CAC34665.1"
/db_xref="GI:13396896"
/translation="MATPHINAEAGDPADVVLMPGDLRAKVIATFLEDAEYNNVR
GMLGPTGYKGRKISVMHGNGIPSCSYTKELJTDGKVIIRVSGCGLVPHVKLR
DVVIGMGACTDSKVNRIKPDHDFALIDFMRNADVAALGIDARVGNLFSADLF
YSPGEMPDVMEKYGILGVEVEAAAGIYGVAAEFGAKALUTICTVSDHIRTHTQTTAAER
QTTFNMIKIALESVLLGDKE"

ORIGIN
Alignment Scores:
Pred. No.: 4,06e-110 Length: 720
Score: 1222.00 Matches: 238
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 99.58% Mismatches: 0
Query Match: 99.76% Indels: 0
DB: 6 Gaps: 0

US-10-035-300A-2 (1-239) x AX087917 (1-720)

Qy 1 MetAlaThrProHisIleAsnAlaGluMetGlyAspPheAlaAspValValLeuMetPro 20
Db 1 ATGCTACCCACACATTAATGACAGAAATGGCGATTTGCTGCTGAGTATGTTGATGCCA 60
Qy 21 GlyAspProLeuArgAlaLysTyrIleAlaGluThrPheLeuGluAspAlaArgGluVal 40
Db 61 GGCAGCCCGCTGCGTGGCAAGTATATTCCTGAACCTTCTTGAAGATGCCCGTGAAGTG 120
Qy 41 AsnAsnValArgGlyMetLeuGlyPheThrGlyThrTyrLysGlyArgLysIleSerVal 60
Db 121 AACACGTTTCGCGGTATGCTGGGCTTACCGGTACTTACAAAGGCCCGCAAAATTTCCGTA 180
```

```
Qy 61 MetGlyHisGlyValGlyIleProSerCysSerIleTyrThrLysGluLeuIleThrAsp 80
Db 181 ATGGGTCAACGGTATGGGTATCCCGTCTCTGCTCCATCTACCAAGAACTGATCACCGAT 240
Qy 81 PheGlyValLysLysIleIleArgValGlySerCysGlyAlaValLeuProHisValLys 100
Db 241 TTCGGCGTGAAGAAATATATCCCGTGGGTTCTGTGGCGAGTTCTCCCGCACATAAA 300
Qy 101 LeuArgAspValValIleGlyMetGlyAlaCysThrAspSerLysValAsnArgIleArg 120
Db 301 CTGGCGGACGTCGTATCGGTATGGGTGCTGCACCGGATTCGAAAGTTTAAACCGATCCGT 360
Qy 121 PheLysAspHisAspPheAlaAlaIleAlaAspPheAspMetValArgAsnAlaValAsp 140
Db 361 TTTAAAGACCATGACTTTGCGCTATCGCTGACTTCGACATGGTGGCTAAACGACGTAGAT 420
Qy 141 AlaLysAlaLeuGlyIleAspAlaArgValGlyAsnLeuPheSerAlaAspLeuPhe 160
Db 421 GCAGCTAAAGCACCTGGGTATTTGATGCTCGCGTGGGTAAACCTGTTCTCCGCTGACCTGTC 480
Qy 161 TyrSerProAspGlyGluMetPheAspValMetGluLysTyrGlyIleLeuGlyValGlu 180
Db 481 TACTCTCCGACGGCGAAATGTTTCACGTGATGGAATAATACGGCATTCCTCGCGGTGAA 540
Qy 181 MetGluAlaAlaGlyIleTyrGlyValAlaAlaGluPheGlyAlaLysAlaLeuThrIle 200
Db 541 ATGGAAGCGGCTGCTATCTACGGCGTCTGCTGAGAAATTTGGCGGAAAGCCCTGACCATC 600
Qy 201 CysThrValSerAspHisIleArgThrHisGluGlnThrThrAlaAlaGluArgGlnThr 220
Db 601 TGCACCGTATCTGACCATCGCACTCAGGAGCAGACCACTGCGCGCTGAGCGTCCAGACT 660
Qy 221 ThrPheAsnAspMetIleLysIleAlaLeuGluSerValLeuLeuGlyAspLysGlu 239
Db 661 ACCTTCAACGACATGATCAAAATCGCACTGGAATCCGTTCTGCTGGCGGATAAAGAG 717

RESULT 4
AX590437 LOCUS AX590437 720 bp DNA linear PAT 27-JAN-2003
DEFINITION Sequence 3 from Patent EP1254959.
ACCESSION AX590437
VERSION AX590437.1 GI:27949070
KEYWORDS
SOURCE Escherichia coli
ORGANISM Escherichia coli
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Escherichia.
REFERENCE 1
AUTHORS Araki, T., Ikeda, I., Matoishi, K., Abe, R., Oikawa, T., Matsuba, Y.,
Ishibashi, H., Nagahara, K. and Fukui, Y.
TITLE Method for producing cytosine nucleoside compounds
JOURNAL Patent: EP 1254959-A 3 06-NOV-2002;
MITSUI CHEMICALS, INC. (JP)
FEATURES
source Location/Qualifiers
1..720
/organism="Escherichia coli"
/mol_type="unassigned DNA"
/db_xref="taxon:562"

ORIGIN
Alignment Scores:
Pred. No.: 4,06e-110 Length: 720
Score: 1222.00 Matches: 238
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 99.58% Mismatches: 0
Query Match: 99.76% Indels: 0
DB: 6 Gaps: 0

US-10-035-300A-2 (1-239) x AX590437 (1-720)

Qy 1 MetAlaThrProHisIleAsnAlaGluMetGlyAspPheAlaAspValValLeuMetPro 20
Db 1 ATGCTACCCACACATTAATGACAGAAATGGCGATTTGCTGCTGAGTATGTTGATGCCA 60
Qy 21 GlyAspProLeuArgAlaLysTyrIleAlaGluThrPheLeuGluAspAlaArgGluVal 40
Db 61 GGCAGCCCGCTGCGTGGCAAGTATATTCCTGAACCTTCTTGAAGATGCCCGTGAAGTG 120
Qy 41 AsnAsnValArgGlyMetLeuGlyPheThrGlyThrTyrLysGlyArgLysIleSerVal 60
Db 121 AACACGTTTCGCGGTATGCTGGGCTTACCGGTACTTACAAAGGCCCGCAAAATTTCCGTA 180
```



```
Db 1 ATGGCTACCCACACATTAATGACAGAAATGGCGAATTCGCTGACGTAGTTTTGTATGCCA 60
Qy 21 GlyAspProLeuArgAlaLysTyrIleAlaGluThrPheLeuGluAspAlaArgGluVal 40
Db 61 GGGACCCGCTGCGTGGAAAGTATTATGCTGAAACTTTCCTTGAAGATGCCCGTGAAGTG 120
Qy 41 AsnAsnValArgGlyMetLeuGlyPheThrGlyThrTyrLysGlyArgLysIleSerVal 60
Db 121 AACACGTTCCGGGTATGCTGGGCTTACCGGTACTTACAAAGCCGCAAAATTCGGTA 180
Qy 61 MetGlyHisGlyValGlyIleProSerCysSerIleTyrThrLysGluLeuIleThrAsp 80
Db 181 ATGGGTCAAGGTATGGGTATCCGCTGCTGCTCCATCTACACCAAGAACTGATCACCGAT 240
Qy 81 PheGlyValLysIleIleArgValGlySerCysGlyAlaValLeuProHisValLys 100
Db 241 TTGGCGGTGAAGAAATAATTATCCGCGTGGGTTCCTGTGGCGAGTTCTGCGCAGCTAATA 300
Qy 101 LeuArgAspValValIleGlyMetGlyValAlaCysThrAspSerLysValAsnArgIleArg 120
Db 301 CTGGCGACGTCGTATCGGTATGGGTGCTGCACCGATTCCAAAGTTAACCGCATCCGT 360
Qy 121 PheLysAspHisAspPheAlaIleAlaAspPheAspMetValArgAsnAlaValAsp 140
Db 361 TTTAAAGACCATGACTTTGCGGTATCGCTGACTTCGACATGGTGGCTAACGCACTAGAT 420
Qy 141 AlaAlaLysAlaLeuGlyIleAspAlaArgValGlyAsnLeuPheSerAlaAspLeuPhe 160
Db 421 GCAGCTAAAGACATGGGTATGTGATGCTGCGGTGGGTAACTGTTCTCCGCTGACCTGTT 480
Qy 161 TyrSerProAspGlyGluMetPheAspValMetGlyLysTyrGlyIleLeuGlyValGlu 180
Db 481 TACTCTCCGACGCGGAAATGTTCCAGCTGATGGAAATACGGCAATTCCTCGCGTGAA 540
Qy 181 MetGluAlaAlaGlyIleTyrGlyValAlaAlaGluPheGlyAlaLysAlaLeuThrIle 200
Db 541 ATGGAAGCGCTGCTATCTACCGGCTGCTGCAAGATTTGCGCGCAAGCCCTGACCATC 600
Qy 201 CysThrValSerAspHisIleArgThrHisGluGlnThrThrAlaAlaGluArgGlnThr 220
Db 601 TGACCGTATCTGACCACTCCGACATCCAGAGACACCTGCGGTGAGCGTGCAGACT 660
Qy 221 ThrPheAsnAspMetLysIleAlaLeuGluSerValLeuLeuGlyAspLysGlu 239
Db 661 ACCTTCAACACATGATCAAAATCGCACTCGAATCCGTTCTGCTGGCGGATAAAGAG 717
```

```
RESULT 5
ECOPNP          1249 bp  DNA  linear  BCT 26-APR-1993
LOCUS          E.coli purine nucleoside phosphorylase (deod) gene, complete cds.
DEFINITION     M60917
ACCESSION     M60917
VERSION       M60917.1 GI:147308
KEYWORDS      purine nucleoside phosphorylase.
SOURCE        Escherichia coli
ORGANISM      Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
              Enterobacteriaceae; Escherichia.
REFERENCE     1 (bases 1 to 1249)
AUTHORS      Hersfield,M.S., Chaffee,S., Koro-Johnson,L., Mary,A., Smith,A.A.
              and Short,S.A.
TITLE         Use of site-directed mutagenesis to enhance the epitope-shielding
              effect of covalent modification of proteins with polyethylene
              glycol
JOURNAL       Proc. Natl. Acad. Sci. U.S.A. 88 (16), 7185-7189 (1991)
MEDLINE      9134430
PUBMED      1714590
COMMENT      source text: Escherichia coli (strain K-12) DNA.
FEATURES     Location/Qualifiers
              1..1249
               /organism="Escherichia coli"
               /mol_type="genomic DNA"
               /strain="K-12"
               /db_xref="taxon:562"
```

```
RBS 111..115
      /note="putative"
gene 123..842
      /genes="deod"
CDS   123..842
      /genes="deod"
      /EC_number="2.4.2.1"
      /note="putative"
      /codon_start=1
      /transl_table=11
      /product="purine nucleoside phosphorylase"
      /protein_id="AAA24401.1"
      /db_xref="GI:147309"
      /translation="MATPHINAEIMGDPADVLMPGDPLRAKYIAETFLBEDAREVNVNR
      GMLFGTGYKGRKISVMGHMGIPSCSIYTKELITDPGVKKLIIRVSCGAVLPHVKLR
      DVVIGAGCTDSKNIRIFKDHDFAAIADFDMWRNVAADAKALGIDARVGNLFSADLF
      YSPDGMFDMERYGILGVMEAAAGIYVAAEFGAKALITCTVSDHRIHTEQTTAAER
      QTTFNDMIKIALESVLLGDKE"
      874..901
      /note="putative"
      /function="terminator for deod transcription"
stem_loop 874..901
      /note="putative"
      /function="terminator for deod transcription"
ORIGIN
Alignment Scores:
Pred. No.:      8,17e-110      Length:      1249
Score:          1222.00        Matches:    238
Percent Similarity: 100.00%    Conservative: 1
Best Local Similarity: 99.58%  Mismatches: 0
Query Match:     99.76%       Indels:     0
DB:              1            Gaps:       0
US-10-035-300A-2 (1-239) x ECOPNP (1-1249)
Qy 1 MetAlaThrProHisIleAsnAlaGluMetGlyAspPheAlaAspValValLeuMetPro 20
Db 123 ATGGCTACCCACACATTAATGCAAAATGGCGCATTCGCTGACGTAGTTTTGTATGCCA 182
Qy 21 GlyAspProLeuArgAlaLysTyrIleAlaGluThrPheLeuGluAspAlaArgGluVal 40
Db 183 GGGACCCGCTGCGTGGAAAGTATTATGCTGAAACTTTCCTTGAAGATGCCCGTGAAGTG 242
Qy 41 AsnAsnValArgGlyMetLeuGlyPheThrGlyThrTyrLysGlyArgLysIleSerVal 60
Db 243 AACACGTTCCGGGTATGCTGGGCTTCACCGGTACTTACAAAGCCGCAAAATTCGTA 302
Qy 61 MetGlyHisGlyValGlyIleProSerCysSerIleTyrThrLysGluLeuIleThrAsp 80
Db 303 ATGGGTCAAGGTATGGGTATCCGCTCCTGCTCATCTACACCAAGAACTGATCACCGAT 362
Qy 81 PheGlyValLysIleIleArgValGlySerCysGlyAlaValLeuProHisValLys 100
Db 363 TTGGCGGTGAAGAAATTAATCCGCTGGGTTCCTGTGGCGCAGTTCTGCCGCACTAATA 422
Qy 101 LeuArgAspValValIleGlyMetGlyAlaCysThrAspSerLysValAsnArgIleArg 120
Db 423 CTGCGCAGCTGCTTATCGGTATGGGTGCTGCACCGATTCCAAAGTTAACCGCATCCGT 482
Qy 121 PheLysAspHisAspPheAlaIleAlaAspPheAspMetValArgAsnAlaValAsp 140
Db 483 TTTAAAGACCATGACTTTCCCGCTATCGGTGACTTCGACATGGTGGGTAACTGTTCTCCGCTGACCTGTT 542
Qy 141 AlaAlaLysAlaLeuGlyIleAspAlaArgValGlyAsnLeuPheSerAlaAspLeuPhe 160
Db 543 GCAGCTAAAGCACTGGGTATTGATGCTCGGTGGGTAACTGTTCTCCGCTGACCTGTT 602
Qy 161 TyrSerProAspGlyGluMetPheAspValMetGlyLysTyrGlyIleLeuGlyValGlu 180
Db 603 TACTCTCCGACGCGCAAAATGTTCCAGCTGATGGAAAAATACGGCATTCCTCGCGGTGAA 662
Qy 181 MetGluAlaAlaGlyIleTyrGlyValAlaAlaGluPheGlyAlaLysAlaLeuThrIle 200
Db 663 ATGGAAGCGGCTGGTATCTACCGGCTGCTGCAAGATTTGGCGCGCAAGCCCTGACCATC 722
```

```

Qy 201 CysThrValSerAspHisIleAArgThrHisGluInThrThrAlaAlaGluArgGlnThr 220
Db 723 TGCACCGGTATCGACCATCGCACTCACGAGCAGACCACTCGCGCTGAGCGTCAGACT 782

Qy 221 ThrPheAsnAspMetIleLysIleAlaLeuGluSerValLeuLeuGluValAspLysGlu 239
Db 783 ACCTTCAACGACATGATCAAAATGCGACTGGAAATCCGTTCTGCTGGCGGATAAAGAG 839

RESULT 6
BD261823
LOCUS BD261823 3031 bp DNA linear PAT 17-JUL-2003
DEFINITION Recombinant bacterial strains for the production of natural
nucleosides and modified analogues thereof.
ACCESSION BD261823.1 GI:33071591
VERSION JP 2002533126-A/12.
KEYWORDS synthetic construct
SOURCE synthetic construct
ORGANISM other sequences; artificial sequences.
REFERENCE 1 (bases 1 to 3031)
AUTHORS Bestetti,G., Cali,S., Ghisotti,D., Orsini,G., Tonon,G. and Zuffi,G.
TITLE Recombinant bacterial strains for the production of natural
nucleosides and modified analogues thereof
JOURNAL Patent: JP 2002533126-A 12 08-OCT-2002;
NORPHARMA SPA
COMMENT OS Artificial Sequence
PN JP 2002533126-A/12
PD 08-OCT-2002
PF 23-DEC-1999 JP 2000591198
PR 23-DEC-1998 IT MI 98A002792
PI GIUSEPPINA BESTETTI, SIMONA CALI, DANIELA GHISOTTI, GAETANO PI
ORSINI,
PI GIANCARLO TONON, GABRIELE ZUFFI
PC C12N15/09, C12N1/21, C12N9/10//C12P19/38, C12P19/40, C12N15/00 CC
Description of Artificial Sequence: udp and deod cloned into CC
PGW746
CC without upstream ptac promoter
FH Key Location/Qualifiers
FT source 1..3031
FT /organism='Artificial Sequence'.

FEATURES
source
1..3031
/organism="synthetic construct"
/mol_type="genomic DNA"
/db_xref="taxon:32630"

ORIGIN
Alignment Scores: 2.52e-109 Length: 3031
Pred. No.: 1222.00 Matches: 238
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 99.58% Mismatches: 0
Query Match: 99.76% Indels: 0
DB: 6 Gaps: 0

US-10-035-300A-2 (1-239) x BD261823 (1-3031)

Qy 1 MetAlaThrProHisIleAsnAlaGluMetGlyAspPheAlaAspValValLeuMetPro 20
Db 101 ATGGCTACCCACACATTAATGACGAATATGGCGAATTCGCTGACGTAGTTTGTATGCCA 160

Qy 21 GlyAspProLeuArgAlaLysIleAlaGluThrPheLeuGluAspAlaArgGluVal 40
Db 161 GGCGACCGCTGCTGCGAAGTATATGCTCAAACTTCTTGAAGATGCCCGTGAAGT 220

Qy 41 AsnAsnValArgGlyMetLeuGlyPheThrGlyThrTyrlsGlyArgLysIleSerVal 60
Db 221 AACAAACGTTCGCGTATGCTGGGCTTACCGGTACTTACAAAGCGCGCAAAATTTCCGTA 280

Qy 61 MetGlyHisGlyValGlyIleProSerCysSerIleTyrlsGlyGluLeuIleThrAsp 80
Db 281 ATGGGTACCGTATGGGTATCCCGTCTGCTCCATCTACCAAGAACTGATCACCAGT 340

```

```

Qy 81 PheGlyValLysLysIleAArgValGlySerCysGlyAlaValLeuProHisValLys 100
Db 341 TTCGCGCTGAAGAAAATTTATCCGCTGGGTTCTCTGTGCGCAGACTTCTCGCGCACGTAATA 400

Qy 101 LeuArgAspValValIleGlyMetGlyAlaCysThrAspSerLysValAsnArgIleArg 120
Db 401 CTGGCGACGCTGTTATCGGTATGGTGCTTCGACCGGATTCAAAAGTTTAAACCGCATCCGT 460

Qy 121 PheLysAspHisAspPheAlaAlaIleAlaAspPheAspMetValArgAsnAlaValAsp 140
Db 461 TTTTAAAGACCATGACTTTTGGCGCTATCGCTGACTTCGACATGGTGGTGAACGACGTAGAT 520

Qy 141 AlaAlaLysAlaLeuGlyIleAspAlaArgValGlyAsnLeuPheSerAlaAspLeuPhe 160
Db 521 GCAGCTAAAGCACCTGGGTATTGATGCTGCGGTGGGTAAACCTGTTCTCCGCTGACCTGTC 580

Qy 161 TyrSerProAspGlyGluMetPheAspValMetGluLysTyrlsGlyIleLeuGlyValGlu 180
Db 581 TACTCTCGGACGGCGGAATGTTTCAGCTGATGGAAAATACGGCATTTCTCGGCGTGGAA 640

Qy 181 MetGluAlaAlaGlyIleTyrlsGlyValAlaAlaGluPheGlyAlaLysAlaLeuThrIle 200
Db 641 ATGGAAGCGGCTGCTATCTACGCGCTGCTGCAGAAATTTGGCGGAAAGCCCTGACCATC 700

Qy 201 CysThrValSerAspHisIleAArgThrHisGluInThrThrAlaAlaGluArgGlnThr 220
Db 701 TGCACCGGTATCTGACCATCCGCACTCACGAGCAGACCACTGCGCTGAGCGTCAGACT 760

Qy 221 ThrPheAsnAspMetIleLysIleAlaLeuGluSerValLeuLeuGluValAspLysGlu 239
Db 761 ACCTTCAACGACATGATCAAAATCGACTGGAAATCCGTTCTGCTGGCGGATAAAGAG 817

RESULT 7
AX027820 3031 bp DNA linear PAT 16-SEP-2000
LOCUS AX027820 Sequence 12 from Patent WO0039307.
ACCESSION AX027820
VERSION AX027820.1 GI:10188664
KEYWORDS synthetic construct
SOURCE synthetic construct
ORGANISM other sequences; artificial sequences.
REFERENCE 1
AUTHORS Bestetti,G., Cali,S., Orsini,G., Tonon,G., Zuffi,G. and Ghisotti,D.
TITLE Recombinant bacterial strains for the production of natural
nucleosides and modified analogues thereof
JOURNAL Patent: WO 0039307-A 12 06-JUL-2000;
BESTETTI GIUSEPPINA (IT); CALI SIMONA (IT); NORPHARMA SPA (IT);
ORSINI GAETANO (IT); TONON GIANCARLO (IT); ZUFFI GABRIELE (IT);
GHISOTTI DANIELA (IT)
FEATURES
source
1..3031
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/notes="udp and deod cloned into pGM746, without upstream
ptac promoter"

ORIGIN
Alignment Scores: 2.52e-109 Length: 3031
Pred. No.: 1222.00 Matches: 238
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 99.58% Mismatches: 0
Query Match: 99.76% Indels: 0
DB: 6 Gaps: 0

US-10-035-300A-2 (1-239) x AX027820 (1-3031)

Qy 1 MetAlaThrProHisIleAsnAlaGluMetGlyAspPheAlaAspValValLeuMetPro 20
Db 101 ATGGCTACCCACACATTAATGACGAATATGGCGAATTCGCTGACGTAGTTTGTATGCCA 160

```


KEYWORDS
SOURCE synthetic construct
ORGANISM synthetic construct
REFERENCE 1 other sequences; artificial sequences.

DEFINITION
TITLE Beateetti.G., Cali.S., Orsini.G., Tonon.G., Zuffi.G. and Ghisotti.D.
AUTHORS Recombinant bacterial strains for the production of natural
nucleosides and modified analogues thereof

JOURNAL
PATENT: WO 0039307-A 13 06-JUL-2000;
BESTETTI GIUSEPPINA (IT); CALI SIMONA (IT); NORPHARMA SPA (IT);
ORSINI GAETANO (IT); TONON GIANCARLO (IT); ZUFFI GABRIELE (IT);
GHISOTTI DANIELA (IT)

FEATURES
source
1..3128
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/note="deod cloned downstream ptac promoter"

ORIGIN
Alignment Scores:
Pred. No.: 2.62e-109 Length: 3128
Score: 1222.00 Matches: 238
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 99.58% Mismatches: 0
Query Match: 99.76% Indels: 0
DB: 6 Gaps: 0

US-10-035-300A-2 (1-239) x AX027821 (1-3128)

Qy 1 MetAlaThrProHisIleAsnAlaGluMetGlyAspPheAlaAspValValLeuMetPro 20
Db 198 ATGGCTACCCACACATTAATGCAGAAATGGCGATTTCGCTGACGTAGTTTGTATGCCA 257
Qy 21 GlyAspProLeuArgAlaLysTyriLeAlaGluThrPheLeuGluAspAlaArgGluVal 40
Db 258 GCGACCGCGTCGCTGGCGAAGTATTGCTGAAACTTTCCTTGAAGATGCCCGTGAAGTG 317
Qy 41 AsnAsnValArgGlyMetLeuGlyPheThrGlyThrTyriLysGlyArgLysIleSerVal 60
Db 318 AACAACTTCGCGTATGCTGGCTTCACCGGTACTTACAAAGCGCGCAAAATTCGGTA 377
Qy 61 MetGlyHisGlyValGlyIleProSerCysSerIleTyriLysGluLeuIleThrAsp 80
Db 378 ATGGGTACCGGTATGGGTATCCGCTCTGCTCCATCTACACCAAGAACTGATCACCAGT 437
Qy 81 PheGlyValLysIleIleArgValGlySerCysGlyAlaValLeuProHisValLys 100
Db 438 TTCGCGTGAAGAAATATATCCGCTGGGTTCCTTGGCGCAGTTCGCCGACGTAAA 497
Qy 101 LeuArgAspValValIleGlyMetGlyAlaCysThrAspSerLysValAsnArgIleArg 120
Db 498 CTGCGCAGCTGCTATCGGTATGGGTATGGTGCCTGCACCGATTCCAAAGTTAACCGCATCGT 557
Qy 121 PheLysAspHisAspPheAlaIleAlaAspPheAspMetValArgAsnAlaValAsp 140
Db 558 TTTAAAGACCATGACTTTTGGCGCTATCGCTGACTTCGACATGTCGCGTAACGCGTAGAT 617
Qy 141 AlaAlaLysAlaLeuGlyIleAspAlaArgValGlyAsnLeuPheSerAlaAspLeuPhe 160
Db 618 GCACCTAAGCAGCTGGGTATGATGCTCGCGTGGGTAACTTCTTCGCTGACCTGTTC 677
Qy 161 TyrSerProAspGlyGluMetPheAspValMetGluLysTyriGlyIleLeuGlyValGlu 180
Db 678 TACTCTCGGACGCGGAAATGTTTCGACGTGATGGAATAATACGCCATTCTCGCGGTGAA 737
Qy 181 MetGluAlaAlaGlyIleTyriGlyValAlaAlaGluPheGlyAlaLysAlaLeuThrIle 200
Db 738 ATGGAAGCGGTGCTATCTACGGGTCTGCTGAGAAATTTGGCGGAAAGCCCTGACCATC 797
Qy 201 CysThrValSerAspHisIleArgThrHisGluGlnThrAlaAlaGluArgGlnThr 220
Db 798 TGCACCGTATCTGACCATCCGCACTCAGCAGCAGCACTGCGCGTGCAGCGTACACT 857

Qy 221 ThrPheAsnAspMetIleLysIleAlaLeuGluSerValLeuLeuGlyAspLysGlu 239
Db 858 ACCTTCAACGACATGATCAAAATCGCAATCCGTTCTGCTGGCGGATAAAGAG 914

RESULT 10
BD261814 3383 bp DNA linear PAT 17-JUL-2003
LOCUS Recombinant bacterial strains for the production of natural
nucleosides and modified analogues thereof.
DEFINITION
TITLE Beateetti.G., Cali.S., Ghisotti.D., Orsini.G., Tonon.G. and Zuffi.G.
AUTHORS Recombinant bacterial strains for the production of natural
nucleosides and modified analogues thereof

JOURNAL
PATENT: JP 2002533126-A 3 08-OCT-2002;
NORPHARMA SPA

COMMENT
OS Artificial Sequence
PN JP 2002533126-A/3
PD 08-OCT-2002
PF 23-DEC-1999 JP 2000591198
PR 23-DEC-1998 IT MI 98A002792
PI GIUSEPPINA BESTETTI, SIMONA CALI, DANIELA GHISOTTI, GAETANO PI
ORSINI,

PI GIANCARLO TONON, GABRIELE ZUFFI
PC C12N15/09, C12N1/21, C12N9/10//C12P19/38, C12P19/40, C12N15/00 CC
Description of Artificial Sequence: Plasmid
CC deod
FH Key Location/Qualifiers
FT gene Location/Qualifiers
source
1..3383
/organism="synthetic construct"
/mol_type="genomic DNA"
/db_xref="taxon:32630"

ORIGIN
Alignment Scores:
Pred. No.: 2.9e-109 Length: 3383
Score: 1222.00 Matches: 238
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 99.58% Mismatches: 0
Query Match: 99.76% Indels: 0
DB: 6 Gaps: 0

US-10-035-300A-2 (1-239) x BD261814 (1-3383)

Qy 1 MetAlaThrProHisIleAsnAlaGluMetGlyAspPheAlaAspValValLeuMetPro 20
Db 240 ATGGCTACCCACACATTAATGCAGAAATGGCGATTTCGCTGACGTAGTTTGTATGCCA 299
Qy 21 GlyAspProLeuArgAlaLysTyriLeAlaGluThrPheLeuGluAspAlaArgGluVal 40
Db 300 GCGACCGCTGCGTGCAGATATATGCTGAAACTTTCCTTGAAGATGCCCGTGAAGTG 359
Qy 41 AsnAsnValArgGlyMetLeuGlyPheThrGlyThrTyriLysGlyArgLysIleSerVal 60
Db 360 AACAACTTCGCGTATGCTGGGTTCACCGGTACTTACAAAGCGCGCAAAATTCGGTA 419
Qy 61 MetGlyHisGlyValGlyIleProSerCysSerIleTyriLysGluLeuIleThrAsp 80
Db 420 ATGGGTACCGTATGGGTATCCCGTCTGCTCCATCTACACCAAGAACTGATCACCAGT 479
Qy 81 PheGlyValLysIleIleArgValGlySerCysGlyAlaValLeuProHisValLys 100
Db 480 TTCGCGTGAAGAAATATATCCGCTGGGTTCCTGTCGCGCAGTTCGCCGACGTAAAA 539
Qy 101 LeuArgAspValValIleGlyMetGlyAlaCysThrAspSerLysValAsnArgIleArg 120

```
Db      540  |||||CTCGCGGACGTGTTATCGGTATGGGTGCTGCACCGATTCCAAAGTTAACCGCATCCGT 599
Qy      121  PheLysAspHisAspPheAlaAlaIleAlaAspPheAspMetValArgAsnAlaValAsp 140
Db      600  TTTTAAAGACCATGACTTTGCGCGTATCGCTGACTTCGACATGGTGGCTAACGCGTAGAT 659
Qy      141  AlaAlaLysAlaLeuGlyIleAspAlaAatqValGlyAsnLeuPheSerAlaAspLeuPhe 160
Db      660  GCAGCTTAAGCACTGGGTATGATGCTCGCGTGGGTAACTGTCTCCGCTGACCTGTTTC 719
Qy      161  TyrSerProAspGlyGluMetPheAspValMetGluLysTyrGlyIleLeuGlyValGlu 180
Db      720  TACTCTCCGCGACGGCGAAATGTTCCGACGTGATGGAAAAATACGGCATTCCTCGCGTGA 779
Qy      181  MetGluAlaAlaGlyIleTyrGlyValAlaAlaGluPheGlyAlaLysAlaLeuThrIle 200
Db      780  ATGGAAGCGGCTGGTATCTACGGCGTGGCTGCAGAAATTTGGCGCGAAGCCCTGACCATC 839
Qy      201  CysThrValSerAspHisIleArgThrHisGluGlnThrThrAlaAlaGluArgGlnThr 220
Db      840  TGCACCGTATCTGACCACATCCGACTTCACGACGACACCACTGCCGCTGAGCGTCAGACT 899
Qy      221  ThrPheAsnAspMetIleLysIleAlaLeuGluSerValLeuLeuGlyAspLysGlu 239
Db      900  ACCTTCAACGACATGATCAAAATCGCACTGGAATCCGTTCTGCTGGCGGATAAAGAG 956

RESULT 11
AX027811 3383 bp DNA linear PAT 16-SEP-2000
LOCUS
DEFINITION Sequence 3 from Patent WO0039307.
ACCESSION AX027811
VERSION AX027811.1 GI:10188655
KEYWORDS
SOURCE
ORGANISM
REFERENCE
1. Bestetti, G., Cali, S., Orsini, G., Tonon, G., Zuffi, G. and Ghisotti, D.
   Recombinant bacterial strains for the production of natural
   nucleosides and modified analogues thereof
   Patent: WO 0039307-A 3 06-JUL-2000;
   BESTETTI GIUSEPPINA (IT); CALI SIMONA (IT); NORPHARMA SPA (IT);
   ORSINI GAETANO (IT); TONON GIANCARLO (IT); ZUFFI GABRIELE (IT);
   GHISOTTI DANIELA (IT)
FEATURES
    source
    Location/Qualifiers
    1..3383
       /organism="synthetic construct"
       /mol_type="unassigned DNA"
       /db_xref="taxon:32630"
       /note="Plasmid"

ORIGIN
Alignment Scores:
Pred. No.: 2.9e-109 Length: 3383
Score: 1222.00 Matches: 238
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 99.58% Mismatches: 0
Query Match: 99.76% Indels: 0
DB: 6 Gaps: 0

US-10-035-300A-2 (1-239) x AX027811 (1-3383)

Qy      1 MetAlaThrProHisIleAsnAlaGluMetGlyAspPheAlaAspValValLeuMetPro 20
Db      240  ATGGCTACCCACACATTAATGCAGAAATGGCGAATTCGCTCAGCTAGTTCGATGGCCA 299
Qy      21 GlyAspProLeuArgAlaLysTyrIleAlaGluThrPheLeuGluAspAlaArgGluVal 40
Db      300  GGGGACCGCGTGGTGGGAATATATGCTGAACCTTCCTTGAGATGCCGCTGAGTG 359
Qy      41 AsnAsnValArgGlyMetLeuGlyPheThrGlyThrTyrLysGlyArgLysIleSerVal 60
```

```
Db      360  AACACAGTTTCGGGTATGCTGGGCTTTCACCGGTACTTACAAAGGCGCAAAATTTCCGTA 419
Qy      61 MetGlyHisGlyValGlyIleProSerCysSerIleTyrThrLysGluLeuIleThrAsp 80
Db      420  ATGGGTACAGGTATGGGTATCCGCTCTGCTCATCTACACCAAGAAGTATGATCCCGAT 479
Qy      81 PheGlyValLysLysIleIleArgValGlySerCysGlyValaValLeuProHisValLys 100
Db      480  TTCGGGCTGAGNAANATTATCCGCTGGGTCTCTGTGGCGCAGTTCTGCGCAGCTAATA 539
Qy      101 LeuArgAspValValIleGlyMetGlyAlaCysThrAspSerLysValAsnArgIleArg 120
Db      540  CTCGCGACGCTGTTATCGGTATGGGTGCTGCACCGATTCCAAAGTTAACCGCATCCGT 599
Qy      121 PheLysAspHisAspPheAlaAlaIleAlaAspPheAspMetValArgAsnAlaValAsp 140
Db      600  TTTTAAAGACCATGACTTTGCCGCTATCGCTGACTTCGACATGGTGGGTAAACCGTAGAT 659
Qy      141 AlaAlaLysAlaLeuGlyIleAspAlaArgValGlyAsnLeuPheSerAlaAspLeuPhe 160
Db      660  GCAGCTTAAGCACTGGGTATGATGCTCGGTGGGTAACTGTCTCCGCTGACCTGTTTC 719
Qy      161 TyrSerProAspGlyGluMetPheAspValMetGluLysTyrGlyIleLeuGlyValGlu 180
Db      720  TACTCTCCGCGACGGCGAAATGTTCCGACGTGATGGAAAAATACGGCATTCCTCGCGTGA 779
Qy      181 MetGluAlaAlaGlyIleTyrGlyValAlaAlaGluPheGlyAlaLysAlaLeuThrIle 200
Db      780  ATGGAAGCGGCTGGTATCTACGGCGTGGCTGCAGAAATTTGGCGCGAAGCCCTGACCATC 839
Qy      201 CysThrValSerAspHisIleArgThrHisGluGlnThrThrAlaAlaGluArgGlnThr 220
Db      840  TGCACCGTATCTGACCACATCCGACTTCACGACGACACCACTGCCGCTGAGCGTCAGACT 899
Qy      221 ThrPheAsnAspMetIleLysIleAlaLeuGluSerValLeuLeuGlyAspLysGlu 239
Db      900  ACCTTCAACGACATGATCAAAATCGCACTGGAATCCGTTCTGCTGGCGGATAAAGAG 956

RESULT 12
BD261825 3934 bp DNA linear PAT 17-JUL-2003
LOCUS
DEFINITION Recombinant bacterial strains for the production of natural
nucleosides and modified analogues thereof.
ACCESSION BD261825
VERSION BD261825.1 GI:33071593
KEYWORDS
SOURCE
ORGANISM
REFERENCE
1. Bestetti, G., Cali, S., Ghisotti, D., Orsini, G., Tonon, G. and Zuffi, G.
   Recombinant bacterial strains for the production of natural
   nucleosides and modified analogues thereof
   Patent: JP 2002533126-A 14 08-OCT-2002;
   NORPHARMA SPA
   OS Artificial Sequence
   PN JP 2002533126-A/14
   PD 08-OCT-2002
   PF 23-DEC-1999 JP 2000591198
   PR 23-DEC-1998 IT MI 98A002792
   PI GIUSEPPINA BESTETTI, SIMONA CALI, DANIELA GHISOTTI, GAETANO PI
   ORSINI,
   PI GIANCARLO TONON, GABRIELE ZUFFI
PC C12N15/09, C12N1/21, C12N9/10//C12P19/38, C12P19/40, C12N15/00 CC
Description of Artificial Sequence: udp and deod cloned CC
downstream ptac
CC promoter
FH Key Location/Qualifiers
FT source 1..3934
FT /organism='Artificial Sequence'.
FT Location/Qualifiers
1..3934
/organism="synthetic construct"
source
```

```
ORIGIN
/mol_type="genomic DNA"
/db_xref="taxon:32630"

Alignment Scores:
Pred. No.: 3,51e-109 Length: 3934
Score: 1222.00 Matches: 238
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 99.58% Mismatches: 0
Query Match: 99.76% Indels: 0
DB: 6 Gaps: 0

US-10-035-300A-2 (1-239) x BD261825 (1-3934)

Qy 1 MetAlaThrProHisIleAsnAlaGluMetGlyAspPheAlaAspValValLeuMetPro 20
Db 198 ATGGCTACCCACACATTAATGCAGAAATGGCGATTTCGCTGACGTAGTTTGTATGCCA 257
Qy 21 GlyAspProLeuArgAlaLysTyrIleAlaGluThrPheLeuGluAspAlaArgGluVal 40
Db 258 GCGCACCCTGCGTGGCGAAGTATATTGCTGAACCTTTCTTGAAGATGCCCGTGAAGTG 317
Qy 41 AsnAsnValArgGlyMetLeuGlyPheThrGlyThrTyrLysGlyValGlyLysIleSerVal 60
Db 318 AACAAAGCTTCGGGTATGCTGGCTTCACCGGTACTTACAAAGGCCGCAAAATTTCCGTA 377
Qy 61 MetGlyHisGlyValGlyIleProSerCysSerIleTyrThrLysGluLeuIleThrAsp 80
Db 378 ATGGCTACGGTATGGGTATCCCGTCTGCTCCATCTACACCAAGAACTGATCACCAGT 437
Qy 81 PheGlyValLysLysIleAlaArgValGlySerCysGlyAlaValLeuProHisValLys 100
Db 438 TTCGCGTGAAGAAATATATCCGCTGGGTTCCTGTGGCGAGTTCTCCCGCACGTAAAA 497
Qy 101 LeuArgAspValIleGlyMetGlyValaCysThrAspSerLysValAsnArgIleArg 120
Db 498 CTGGCGACGTCGTATTACGGTATGGGTGCTGCACCGGATTCCAAAGTTAACCGCATCCGT 557
Qy 121 PheLysAspHisAspPheAlaAlaIleAlaAspPheAspMetValArgAsnAlaValAsp 140
Db 558 TTTAAAGACCATGACTTTGCCGCTATCGCTGACATTCGACATGGTGGTAAACGCAGTAGAT 617
Qy 141 AlaAlaLysAlaLeuGlyIleAspAlaArgValGlyAsnLeuPheSerAlaAspLeuPhe 160
Db 618 GCAGCTAAAGCACCTGGGTATTGATGCTCGCTGGGTAACTTCTTCGCGCTGACCTGTTTC 677
Qy 161 TyrSerProAspGlyGluMetPheAspValMetGluLysTyrGlyIleLeuGlyValGlu 180
Db 678 TACTCTCCGGACGGCGAAATGTTCCAGCTGATGGAAATACGGCAATTCCTCGCGGTGGAA 737
Qy 181 MetGluAlaAlaGlyIleTyrGlyValAlaAlaGluPheGlyAlaLysAlaLeuThrIle 200
Db 738 ATGGAAAGCGCTGGTATCTACGGGCTCGCTGAGAAATTTGGCGGAAAGCCCTGACCATC 797
Qy 201 CysThrValSerAspHisIleArgThrHisGluGlnThrThrAlaAlaGluArgGlnThr 220
Db 798 TGCACCGTATCTGACCATCCGCACTCACGAGCAGACCACTGCGCGCTGAGCGTCAGACT 857
Qy 221 ThrPheAsnAspMetIleLysIleAlaLeuGluSerValLeuLeuGlyAspLysGlu 239
Db 858 ACCTTCAACGACATGATCAAAATCGCACTGGAAATCCGTTCTGCTGGCGGATAAAGAG 914

RESULT 13
AX027822 3934 bp DNA linear PAT 16-SEP-2000
LOCUS
DEFINITION Sequence 14 from Patent W00039307.
ACCESSION AX027822
VERSION AX027822.1 GI:10188666
KEYWORDS
SOURCE synthetic construct
ORGANISM synthetic construct
other sequences; artificial sequences.
REFERENCE 1
```

```
AUTHORS Bestetti,G., Cali,S., Orsini,G., Tonon,G., Zuffi,G. and Ghiocetti,D.
TITLE Recombinant bacterial strains for the production of natural
JOURNAL nucleosides and modified analogues thereof
Patent: WO 0039307-A 14 06-JUL-2000;
BESTETTI GIUSEPPINA (IT); CALI SIMONA (IT); NORPHARMA SPA (IT);
ORSINI GAETANO (IT); TONON GIANCARLO (IT); ZUFFI GABRIELE (IT);
GHISOFTI DANIELA (IT)
FEATURES
source Location/Qualifiers
1..3934
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/notes="udp and deod cloned downstream ptac promoter"
```

```
ORIGIN
Alignment Scores:
Pred. No.: 3,51e-109 Length: 3934
Score: 1222.00 Matches: 238
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 99.58% Mismatches: 0
Query Match: 99.76% Indels: 0
DB: 6 Gaps: 0
```

```
US-10-035-300A-2 (1-239) x AX027822 (1-3934)

Qy 1 MetAlaThrProHisIleAsnAlaGluMetGlyAspPheAlaAspValValLeuMetPro 20
Db 198 ATGGCTACCCACACATTAATGCAGAAATGGCGATTTCGCTGACGTAGTTTGTATGCCA 257
Qy 21 GlyAspProLeuArgAlaLysTyrIleAlaGluThrPheLeuGluAspAlaArgGluVal 40
Db 258 GCGCACCCTGCGTGGCGAAGTATATTGCTGAACCTTTCTTGAAGATGCCCGTGAAGTG 317
Qy 41 AsnAsnValArgGlyMetLeuGlyPheThrGlyThrTyrLysGlyValGlyLysIleSerVal 60
Db 318 AACAAAGCTTCGGGTATGCTGGGTTCACCGGTACTTACAAAGGCCGCAAAATTTCCGTA 377
Qy 61 MetGlyHisGlyValGlyIleProSerCysSerIleTyrThrLysGluLeuIleThrAsp 80
Db 378 ATGGCTACGGTATGGGTATCCCGTCTGCTCCATCTACACCAAGAACTGATCACCAGT 437
Qy 81 PheGlyValLysLysIleAlaArgValGlySerCysGlyAlaValLeuProHisValLys 100
Db 438 TTCGCGTGAAGAAATATATCCGCTGGGTTCCTGTGGCGAGTTCTCCCGCACGTAAAA 497
Qy 101 LeuArgAspValIleGlyMetGlyValaCysThrAspSerLysValAsnArgIleArg 120
Db 498 CTGGCGACGTCGTATTACGGTATGGGTGCTGCACCGGATTCCAAAGTTAACCGCATCCGT 557
Qy 121 PheLysAspHisAspPheAlaAlaIleAlaAspPheAspMetValArgAsnAlaValAsp 140
Db 558 TTTAAAGACCATGACTTTGCCGCTATCGCTGACATTCGACATGGTGGTAAACGCAGTAGAT 617
Qy 141 AlaAlaLysAlaLeuGlyIleAspAlaArgValGlyAsnLeuPheSerAlaAspLeuPhe 160
Db 618 GCAGCTAAAGCACCTGGGTATTGATGCTCGCTGGGTAACTTCTTCGCGCTGACCTGTTTC 677
Qy 161 TyrSerProAspGlyGluMetPheAspValMetGluLysTyrGlyIleLeuGlyValGlu 180
Db 678 TACTCTCCGGACGGCGAAATGTTCCAGCTGATGGAAATACGGCAATTCCTCGCGGTGGAA 737
Qy 181 MetGluAlaAlaGlyIleTyrGlyValAlaAlaGluPheGlyAlaLysAlaLeuThrIle 200
Db 738 ATGGAAAGCGCTGGTATCTACGGGCTCGCTGAGAAATTTGGCGGAAAGCCCTGACCATC 797
Qy 201 CysThrValSerAspHisIleArgThrHisGluGlnThrThrAlaAlaGluArgGlnThr 220
Db 798 TGCACCGTATCTGACCATCCGCACTCACGAGCAGACCACTGCGCGCTGAGCGTCAGACT 857
Qy 221 ThrPheAsnAspMetIleLysIleAlaLeuGluSerValLeuLeuGlyAspLysGlu 239
Db 858 ACCTTCAACGACATGATCAAAATCGCACTGGAAATCCGTTCTGCTGGCGGATAAAGAG 914
```



```

RESULT 14
BD261816          4189 bp      DNA      linear      PAT 17-JUL-2003
LOCUS
DEFINITION
Recombinant bacterial strains for the production of natural
nucleosides and modified analogues thereof.
ACCESSION
BD261816
VERSION
BD261816.1 GI:33071584
KEYWORDS
JP 2002533126-A/5.
SOURCE
synthetic construct
ORGANISM
other sequences; artificial sequences.
REFERENCE
1 (bases 1 to 4189)
AUTHORS
Bestetti,G., Cali,S., Ghisotti,D., Orsini,G., Tonon,G. and Zuffi,G.
TITLE
Recombinant bacterial strains for the production of natural
nucleosides and modified analogues thereof
JOURNAL
Patent: JP 2002533126-A 5 08-OCT-2002;
NORPHARMA SPA
COMMENT
OS Artificial Sequence
PN JP 2002533126-A/5
PD 08-OCT-2002
PF 23-DEC-1999 JP 2000591198
PR 23-DEC-1998 IT MI 98A002792
PI GIUSEPPINA BESTETTI, SIMONA CALI, DANIELA GHISOTTI, GAETANO PI
ORSINI,
PI GIANCARLO TONON, GABRIELE ZUFFI
PC C12N15/09, C12N1/21, C12N9/10//C12P19/38, C12P19/40, C12N15/00 CC
Description of Artificial Sequence: Plasmid
CC udp
CC deod
FH Key Location/Qualifiers
FT gene (243)..(1021)
FT gene (1037)..(1766).
FEATURES
source
1..4189
/organism="synthetic construct"
/mol_type="genomic DNA"
/db_xref="taxon:32630"
ORIGIN
Alignment Scores:
Pred. No.: 3.8e-109 Length: 4189
Score: 1222.00 Matches: 238
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 99.58% Mismatches: 0
Query Match: 99.76% Indels: 0
DB: 6 Gaps: 0
US-10-035-300A-2 (1-239) x BD261816 (1-4189)
Qy 1 MetAlaThrProHisIleAsnAlaGluMetGlyAspPheAlaAspValValLeuMetPro 20
Db 1046 ATGGCTACCCACACATTAAATGCAGAAATGGCGATTTCGCTGACGTAGTTTGTATGCCA 1105
Qy 21 GlyAspProLeuArgAlaLysTyrIleAlaGluThrPheLeuGluAspAlaArgGluVal 40
Db 1106 GCGACCGCTGCGTGCAGAGTATATTGCTGAAACTTTCCTTGAAGATGCCGGAAGTG 1165
Qy 41 AsnAsnValArgGlyMetLeuGlyPheThrGlyThrTyrLysGlyArgLysIleSerVal 60
Db 1166 AACAACTTCGCGGTATGCTGGGCTTCACCGTACTTACAAAGCCCAAAATTTCCGTA 1225
Qy 61 MetGlyHisGlyValGlyIleProSerCysSerIleTyrThrLysGluLeuIleThrAsp 80
Db 1226 ATGGGTACCGGTATGGGTATCCCGTCTGCTCCATCTACACCAAGACGTATCACCAGT 1285
Qy 81 PheGlyValLysTyrIleArgValGlySerCysGlyAlaValLeuProHisValLys 100
Db 1286 TTCGGCTGAAGAAATATATCCGCTGGGTTCTGTGGCGAGTCTGCGCGACGTAAAA 1345
Qy 101 LeuArgAspValValIleGlyMetGlyAlaCysThrAspSerLysValAsnArgIleArg 120
Db 1346 CTGGCGACGCTGTTATCGGTATGGGTGCTGCACCGCATTCCTCAAAAGTTAACCCGATCCGT 1405

```

```

Qy 121 PheLysAspHisAspPheAlaAlaIleAlaAspPheAspMetValArgAsnAlaValAsp 140
Db 1406 TTTAAGACCATGACATTGCGGCTATCGCTGACTTCGACATGTCGCTAACGCAGTAGAT 1465
Qy 141 AlaAlaLysAlaLeuGlyIleAspAlaArgValGlyAsnLeuPheSerAlaAspLeuPhe 160
Db 1466 GCAGCTAAAGCACTGGGTATTGATGCTCGCGTGGTAACCTGTTCTCCGCTGACCTGTTTC 1525
Qy 161 TyrSerProAspGlyGluMetPheAspValMetGlyLysTyrGlyIleLeuGlyValGlu 180
Db 1526 TACTCTCCGACGCGGAAATGTTGACGCTGATGGAATAATACGCGATTCTCGCGGTGAA 1585
Qy 181 MetGluAlaAlaGlyIleTyrGlyValAlaAlaGluPheGlyAlaLysAlaLeuThrIle 200
Db 1586 ATGGAAGCGCTGGTATCTACGGGCTCGCTGCAGAAATTTGGCGGAAAGCCCTGACCATC 1645
Qy 201 CysThrValSerAspHisIleArgThrHisGluGlnThrThrAlaAlaGluArgGlnThr 220
Db 1646 TGCACCGTATCTGACCACATCCGCACTTCACGAGCAGACCACTCCGCTGAGCGTCAGACT 1705
Qy 221 ThrPheAsnAspMetIleLysIleAlaLeuGluSerValLeuLeuGlyAspLysGlu 239
Db 1706 ACCTTCAACGACATGATCAAAATCGCACTGGAATCCGTTCTGCTGGCGGATAAAGAG 1762
RESULT 15
AX027813          4189 bp      DNA      linear      PAT 16-SEP-2000
LOCUS
DEFINITION
Sequence 5 from Patent WO0039307.
ACCESSION
AX027813
VERSION
AX027813.1 GI:10188657
KEYWORDS
synthetic construct
SOURCE
synthetic construct
ORGANISM
other sequences; artificial sequences.
REFERENCE
1
AUTHORS
Bestetti,G., Cali,S., Orsini,G., Tonon,G., Zuffi,G. and Ghisotti,D.
TITLE
Recombinant bacterial strains for the production of natural
nucleosides and modified analogues thereof
JOURNAL
Patent: WO 0039307-A 5 06-JUL-2000;
BESTETTI GIUSEPPINA (IT) ; CALI SIMONA (IT) ; NORPHARMA SPA (IT) ;
ORSINI GAETANO (IT) ; TONON GIANCARLO (IT) ; ZUFFI GABRIELE (IT) ;
GHISOTTI DANIELA (IT)
FEATURES
Location/Qualifiers
source
1..4189
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/notes="Plasmid"
ORIGIN
Alignment Scores:
Pred. No.: 3.8e-109 Length: 4189
Score: 1222.00 Matches: 238
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 99.58% Mismatches: 0
Query Match: 99.76% Indels: 0
DB: 6 Gaps: 0
US-10-035-300A-2 (1-239) x AX027813 (1-4189)
Qy 1 MetAlaThrProHisIleAsnAlaGluMetGlyAspPheAlaAspValValLeuMetPro 20
Db 1046 ATGGCTACCCACACATTAAATGCAGAAATGGCGATTTCGCTGACGTAGTTTGTATGCCA 1105
Qy 21 GlyAspProLeuArgAlaLysTyrIleAlaGluThrPheLeuGluAspAlaArgGluVal 40
Db 1106 GCGACCGCTGCGTGCAGAGTATATTGCTGAAACTTTCCTTGAAGATGCCGGAAGTG 1165
Qy 41 AsnAsnValArgGlyMetLeuGlyPheThrGlyThrTyrLysGlyArgLysIleSerVal 60
Db 1166 AACAACTTCGCGGTATGCTGGGCTTCACCGTACTTACAAAGCCCAAAATTTCCGTA 1225
Qy 61 MetGlyHisGlyValGlyIleProSerCysSerIleTyrThrLysGluLeuIleThrAsp 80
Db 1226 ATGGGTACCGGTATGGGTATCCCGTCTGCTCCATCTACACCAAGACGTATCACCAGT 1285
Qy 81 PheGlyValLysTyrIleArgValGlySerCysGlyAlaValLeuProHisValLys 100
Db 1286 TTCGGCTGAAGAAATATATCCGCTGGGTTCTGTGGCGAGTCTGCGCGACGTAAAA 1345
Qy 101 LeuArgAspValValIleGlyMetGlyAlaCysThrAspSerLysValAsnArgIleArg 120
Db 1346 CTGGCGACGCTGTTATCGGTATGGGTGCTGCACCGCATTCCTCAAAAGTTAACCCGATCCGT 1405

```



```
|||||:|||||
Db 1226 ATGGGTCAAGGTATCCCGTCTCCATCTACCAAAAGAACTGATCACCGAT 1285
Qy 81 PheGlyValLysLysIleIleArgValGlySerCysGlyAlaValLeuProHisValLys 100
Db 1286 TTCGGCGTGAAGAAATATCCGCGTGGGTTCCTGTGGCGCAGTTCTCCCGCAGCTAAAA 1345
Qy 101 LeuArgAspValValIleGlyMetGlyAlaCysThrAspSerLysValAsnArgIleArg 120
Db 1346 CTGCGCGACGTCGTATCGGTATGGTGCTGCACCGATTCCAAAGTTAACCGCATCCGT 1405
Qy 121 PheLysAspHisAspPheAlaAlaIleAlaAspPheAspMetValArgAsnAlaValAsp 140
Db 1406 TTAAAGACCATGACTTTGCCGCTATCGCTGACTTCGACATGGTGCCTAACGCGATAGAT 1465
Qy 141 AlaAlaLysAlaLeuGlyIleAspAlaArgValGlyAsnLeuPheSerAlaAspLeuPhe 160
Db 1466 GCAGCTAAGCACCTGGGTATTGATGCTCGCGTGGGTAACTGTCTCCGCTGACCTGTTT 1525
Qy 161 TyrSerProAspGlyGluMetPheAspValMetGluLysTyrGlyIleLeuGlyValGlu 180
Db 1526 TACTCTCCGCGACGGCGAAATGTTGACGTGATGGAAAAATACGGCATTCTCGGCGTGAA 1585
Qy 181 MetGluAlaAlaGlyIleTyrGlyValAlaAlaGluPheGlyAlaLysAlaLeuThrIle 200
Db 1586 ATGGAAGCGGCTGGTATCTACGGCGTCTGCTGCAGAAATTTGGCGGAAAGCCCTGACCATC 1645
Qy 201 CysThrValSerAspHisIleArgThrHisGluGlnThrThrAlaAlaGluArgGlnThr 220
Db 1646 TGCACCGTATCTGACCACATCGGCACCTCAGAGCAGACCACTGCCGCTGAGCGTCAGACT 1705
Qy 221 ThrPheAsnAspMetIleLysIleAlaLeuGluSerValLeuLeuGlyAspLysGlu 239
Db 1706 ACCTTCAACGACATGATCAAAATCGCACTGGAATCCGTTCTGCTGGCGGATAAAGAG 1762
```

Search completed: July 23, 2005, 03:20:45
Job time : 5344 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: July 22, 2005, 22:27:52 ; Search time 677 Seconds
(without alignments)
2089.835 Million cell updates/sec

Title: US-10-035-300A-2

Perfect score: 1225

Sequence: 1 MATPHINAEAGDFADVVLMP.....TTFNDMTKIALESVLLGDKE 239

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool_p/US1003300/runat_18072005_165350_27964/app_query.fasta_1.391
-DB=N_Geneseq_16Dec04 -QFWT=fastcap -SUFFIX=ring -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blotsum62 -TRANS=human40.cdi
-LIST=45 -DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFWT=pct -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US1003300.scgn.1.1.708 @runat_18072005_165350_27964 -NCPU=6 -ICPU=3
-NO MMAP -LARGQUERY -NEG SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

N_Geneseq_16Dec04:*
1: Geneseq1980s:*
2: Geneseq1990s:*
3: Geneseq2000s:*
4: Geneseq2001as:*
5: Geneseq2001bs:*
6: Geneseq2002as:*
7: Geneseq2002bs:*
8: Geneseq2003as:*
9: Geneseq2003bs:*
10: Geneseq2003cs:*
11: Geneseq2003ds:*
12: Geneseq2004as:*
13: Geneseq2004bs:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1225	100.0	720	8 AAD56237	Aad56237 Escherich
2	1222	99.8	720	5 AAF55440	Aaf55440 Nucleotid
3	1222	99.8	720	8 AAD54112	Aad54112 Escherich
4	1222	99.8	720	12 ADL27841	Adl27841 E coli de
5	1222	99.8	752	6 ABL51595	Ab151595 Purine nu

6	1222	99.8	3383	3	AAA51632	Aaa51632 Plaemid p
7	1222	99.8	4189	3	AAA51634	Aaa51634 Plaemid p
8	1222	99.8	5013	9	ACA63355	ACA63355 E. coli D
9	1222	99.8	5013	10	AAD59423	Aad59423 E.coli pu
10	1222	99.8	5013	10	AD86127	Ad86127 E. coli p
11	1222	99.8	5495	3	AAA51633	Aaa51633 Plaemid p
12	1222	99.8	6301	3	AAA51635	Aaa51635 Plaemid p
13	1218	99.4	720	8	AAD56238	Aad56238 Escherich
14	1213	99.0	720	5	AAF55446	Aaf55446 Nucleotid
15	1162	94.9	732	11	ACH95836	Ach95836 Klebsiell
16	1104	90.1	3066	3	AAA51641	Aaa51641 PGM747 co
17	1103	90.0	3165	3	AAA51642	Aaa51642 PGM751 co
18	1103	90.0	3979	3	AAA51643	Aaa51643 PGM800 wi
19	1103	90.0	6117	3	AAA51644	Aaa51644 PGM807 wi
20	1101	89.9	5303	3	AAA51636	Aaa51636 Plaemid p
21	1092	89.1	6373	3	AAA51639	Aaa51639 PGM795 co
22	1087	88.7	6343	3	AAA51638	Aaa51638 Plaemid p
23	1067	87.1	717	10	ACF69888	Acf69888 Photorhab
24	1067	87.1	110000	10	ACF65385_0	Acf65385 Photorhab
25	1067	87.1	110000	10	ACF67367_29	Continuation (30 o
26	1057	86.3	795	10	ADF02086	Adf02086 Bacterial
27	1017	83.0	6701	13	ADT05517	Adt05517 Haemophil
28	1017	83.0	85814	13	ADT05644	Adt05644 Haemophil
29	1013	82.7	110000	2	AAT42063_05	Continuation (6 of
30	930.5	76.0	2807	5	AAS86794	Aas86794 DNA encod
31	912	74.4	1893	5	AAS88507	Aas88507 DNA encod
32	871.5	71.1	1112	5	AAS89111	Aas89111 DNA encod
33	778	63.5	110000	6	ABA92787_5	Continuation (6 of
34	700.5	57.2	783	5	AAS86792_	Aas86792 DNA encod
35	693.5	56.6	899	13	ADR12667	Adr12667 Bacillus
36	675	55.1	843	2	AAK14178	Aak14178 H. pylori
37	671.5	54.8	110000	6	ABQ69245_19	Continuation (20 o
38	671.5	54.8	319630	6	ABQ67194_	Abq67194 Listeria
39	670.5	54.7	1171	6	ABQ70303	Abq70303 Listeria
40	670.5	54.7	110000	6	ABR03041_19	Continuation (20 o
41	659.5	53.8	2040	2	AAT77102	Aat77102 DNA encod
42	659	53.8	110000	6	ABA90521_09	Continuation (10 o
43	635	51.8	747	9	ADB09706_	Adb09706 Alloiococ
44	635	51.8	110000	9	ADB12064_10	Continuation (11 o
45	607	49.6	735	10	ADC91558	Adc91558 E. faeciu

ALIGNMENTS

RESULT 1	
AAD56237	
ID	AAD56237 standard; DNA; 720 BP.
XX	
AC	AAD56237;
XX	
DT	07-AUG-2003 (first entry)
XX	
DE	Escherichia coli mutant PNP enzyme (M65V) encoding DNA #SEQ ID 1.
XX	
KW	Purine nucleoside phosphorylase; mutant; tumour; PNP; enzyme; cancer;
KW	leukaemia; glioblastoma; lymphoma; melanoma; gene therapy;
KW	gene; ds.
XX	
OS	Escherichia coli.
OS	Synthetic.
XX	
FH	Key
FT	CDS
FT	Location/Qualifiers
FT	1..720
FT	/*tag= a
FT	/product= "Mutant PNP enzyme"
XX	
PN	WO2003035012-A2.
XX	
PD	01-MAY-2003.
XX	
PF	28-OCT-2002; 2002WO-US034626.
XX	
PR	26-OCT-2001; 2001US-00035300.

XX (UABR-) UAB RES FOUND.
 PA (SOUR) SOUTHERN RES INST.
 PA (CORR) CORNELL RES FOUND INC.
 XX
 PI Ealick SE, Parker WB, Seerist JA, Sorscher EJ;
 XX
 DR WPI: 2003-421350/39.
 DR P-PSDB; AAE37217.
 XX
 PT Novel purified mutant purine cleaving enzyme having different biological
 PT activity than a wild-type purine cleaving enzyme, useful for treating a
 PT pathological condition characterized by abnormal cell growth.
 XX
 PS Claim 10; Page 63-64; 67pp; English.
 XX
 CC The invention relates to mutant purine nucleoside phosphorylase (PNP)
 CC enzymes and nucleic acid molecules encoding such enzymes. These mutant
 CC enzymes have greater biological activity than wild-type enzymes. PNP
 CC enzymes are useful for treating a pathological condition characterised by
 CC abnormal cell growth such as that occurring in cancer of the skin,
 CC breast, bone, colon, head or neck, kidney, larynx, lung, liver, ovary,
 CC nasopharynx, oesophagus, pancreas, prostate, bladder, rectum, stomach,
 CC thyroid, testicle and uterus as well as other conditions characterised by
 CC abnormal cell growth such as myeloid leukaemia, glioblastoma and
 CC lymphoma. Sequences of the invention are useful for treating metastatic
 CC solid tumours such as melanoma, pancreatic, liver or colonic carcinoma or
 CC to activate prodrugs in the treatment of cancer. They are useful for
 CC treating virally infected cells. They are also useful in gene therapy.
 CC The present sequence is Escherichia coli mutant PNP enzyme (Me5V)
 CC encoding DNA
 XX
 SQ Sequence 720 BP; 171 A; 130 C; 187 G; 172 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 1.05e-139 Length: 720
 Score: 1225.00 Matches: 239
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 8 Gaps: 0

US-10-035-300A-2 (1-239) x AAD56237 (1-720)

QY 1 MetAlaThrProHisIleAenAlaGluMetGlyAspPheAlaAspValValLeuMetPro 20
 DB 1 ATGGCTACCCACACATTAATGACAGAAATGGCGATTTCGCTGACGTAGTTTGTATGCCA 60
 QY 21 GlyAspProLeuArgAlaLysTyrIleAlaGluThrPheLeuGluAspAlaArgGluVal 40
 DB 61 GCGACCCGCTGCTGCGAAGTATATTGCTGAAACTTCTTGAAGATGCCGTGAAGTG 120
 QY 41 AsnAsnValArgGlyMetLeuGlyPheThrGlyThrTyrLysGlyArgLysIleSerVal 60
 DB 121 AACAACTTCGCGGTATGCTGGGCTTCACCGGTACTTACAAAGCCGCCAAAATTCGCGTA 180
 QY 61 MetGlyHisGlyValGlyIleProSerCysSerIleTyrThrLysGluLeuIleThrAsp 80
 DB 181 ATGGGTCAACGGTGTGTATATCCCGTCTCTGCTCCATCTACACCAAGAACTGATCACCAGT 240
 QY 81 PheGlyValLysIleIleArgValGlySerCysGlyAlaValLeuProHisValLys 100
 DB 241 TTCGGCTGAAGAAATATATCCGGTGGTTCCTGTGCGCAGTTCTGCGCAGTAATAA 300
 QY 101 LeuArgAspValValIleGlyMetGlyAlaCysThrAspSerLysValAsnArgIleArg 120
 DB 301 CTGCGCACGTGCTATCGGTATGGTGGTGCCTGCACCGATTCCTCAAGTAAACCGCATCCGT 360
 QY 121 PheLysAspHisAspPheAlaAlaIleAlaAspPheAspMetValArgAsnAlaValAsp 140
 DB 361 TTTAAACACCATGACTTTGGCGCTATCGCTGACTTCGACATGTGTGCGTAACGCAATGAT 420
 QY 141 AlaAlaLysAlaLeuGlyIleAspAlaArgValGlyAsnLeuPheSerAlaAspLeuPhe 160

DB 421 GCAGCTAAAGCACTGGGTATTGATGCTCGGTGGGTAACTGTTCTCGCTGACCTGTC 480
 QY 161 TyrSerProAspGlyGluMetPheAspValMetGluLysTyrGlyIleLeuGlyValGlu 180
 DB 481 TACTCTCCGACGCGGAAATGTTTCGACGTGATGAAAAATACGGCATTCCTCGGCGTGAA 540
 QY 181 MetGluAlaAlaGlyIleTyrGlyValAlaAlaGluPheGlyAlaLysAlaLeuThrIle 200
 DB 541 ATGAGACGGCTGGTATCTACGGCTGCTGCAAGATTTCGCGGAAAGCCCTGACCATC 600
 QY 201 CysThrValSerAspHisIleArgThrHisGluGlnThrThrAlaAlaGluArgGlnThr 220
 DB 601 TGCACCGTATCTGACCACATCCGCACTCACGAGCAGACCACTGCGCGTGGAGCGTCAGACT 660
 QY 221 ThrPheAsnAspMetIleLysIleAlaLeuGluSerValLeuLeuGlyAspLysGlu 239
 DB 661 ACCTTCAACGACATGATCAAAATCGCACTGGAATCCGTTCTGTGGCGGATAAAGAG 717
 RESULT 2
 AAF55440
 ID AAF55440 standard; DNA; 720 BP.
 XX
 AC AAF55440;
 XX
 DT 29-MAY-2001 (first entry)
 XX
 DE Nucleotide sequence of a purine nucleoside phosphorylase.
 XX
 KW Deoxyribonucleoside; deoxyribose 1-phosphate; thymidine phosphorylase;
 KW purine nucleoside phosphorylase; phosphopentose mutase;
 KW phosphopentose aldolase; fructose 1,6-diphosphate aldolase;
 KW deoxyribokinase; nucleoside 2-deoxyriboseyltransferase; ss.
 XX
 OS Escherichia coli.
 XX
 FH Key Location/Qualifiers
 FT CDS 1..720
 FT /tag= a
 FT /product= "purine nucleoside phosphorylase"
 XX
 WO200114566-A2.
 XX
 PD 01-MAR-2001.
 XX
 PF 18-AUG-2000; 2000WO-EP008088.
 XX
 PR 20-AUG-1999; 99EP-00116425.
 XX
 PA (HOFF) ROCHE DIAGNOSTICS GMBH.
 PA (INSP) INST PASTEUR.
 PA (PHAR-) PHARMA-WALDHOF GMBH & CO KG.
 XX
 PI Tischer W, Ihlenfeldt H, Barzu O, Sakamoto H, Pistotnik E;
 PI Marliere P, Pochet S;
 XX
 WPI: 2001-235026/24.
 P-PSDB; AAB67584.
 XX
 PT In vitro enzymatic synthesis of deoxyribonucleosides comprises reacting
 PT deoxyribose 1-phosphate and a nucleobase to form a deoxyribonucleoside
 PT and an inorganic phosphate.
 XX
 PS Disclosure; Page 44-45; 73pp; English.
 XX
 CC The present sequence encodes a purine nucleoside phosphorylase enzyme.
 CC This enzyme is involved in the biosynthesis of deoxyribonucleosides, and
 CC is used in the method of the invention. The specification describes a
 CC method for the in vitro enzymatic synthesis of deoxyribonucleosides. The
 CC method comprises reacting deoxyribose 1-phosphate and a nucleobase to
 CC form a deoxyribonucleoside and an inorganic phosphate. Enzymes which may
 CC be used in the method of the invention include thymidine phosphorylase,
 CC purine nucleoside phosphorylase, phosphopentose mutase, phosphopentose

Db 301 CTGGCGACGCTGTTATCGGTATGGGTGCTGCACCGGATTCCAAAGTTAAACCGCATCCGT 360
Qy 121 PheLysAspHisAspPheAlaAlaIleAlaAspPheAspMetValArgAsnAlaValAsp 140
Db 361 TTTAAAGACCATGACTTTGGCGCTATCGCTGACTTTCGACATGGTGGGTAAACGCGATAGAT 420
Qy 141 AlaAlaLysAlaLeuGlyIleAspAlaArgValGlyAsnLeuPheSerAlaAspLeuPhe 160
Db 421 GCAGCTAAAGCATGGGTATGTATGCTCGGTGGGTAACTGTTCTCCGTGACCTGTTTC 480
Qy 161 TyrSerProAspGlyGluMetPheAspValMetGluLysTyrGlyIleLeuGlyValGlu 180
Db 481 TACTCTCCGACGCGGAAATGTTTCGACGTGATGCAAAATACGGCATTTCTCGCGGTGAA 540
Qy 181 MetGluAlaAlaGlyIleTyrGlyValAlaAlaGluPheGlyAlaLysAlaLeuThrIle 200
Db 541 ATGGAACGGCTGGTATCTACGGCGTGCCTGCAGAAATTTGGCGGAAAGCCCTGACCATC 600
Qy 201 CysThrValSerAspHisIleArgThrHisGluGlnThrThrAlaAlaGluArgGlnThr 220
Db 601 TGCACCGTATCTGACCATCTCCGACTCAGAGCGAGACCACTGCGCGTGAAGCTCAGACT 660
Qy 221 ThrPheAsnAspMetIleLysIleAlaLeuGluSerValLeuLeuGlyAspLysGlu 239
Db 661 ACCTTCAACGACATGATCAAAATCGCACTGGAATCCGTTCTGCTGGCGGATAAAGAG 717
RESULT 4
ADL27841
ID ADL27841 standard; DNA; 720 BP.
XX
AC ADL27841;
XX
DT 20-MAY-2004 (first entry)
XX
DE E coli deoD coding sequence.
XX
KW ds; gene; selection method; yield; pathogen resistance;
KW nutritional quality; stress resistance; plant; transgenic.
XX
OS Escherichia coli.
XX
PN WO2004013333-A2.
XX
PD 12-FEB-2004.
XX
PF 18-JUL-2003; 2003WO-EP007877.
XX
PR 26-JUL-2002; 2002DE-01034287.
XX
PA (BADI) BASF PLANT SCI GMBH.
XX
PI Kock M, Frank M, Badur R;
XX
DR WPI; 2004-157134/15.
XX
DR P-PSDB; ADL27842.
XX
PT Method for transforming plant cells, useful for preparing transgenic
PT plants for e.g. food or production of chemicals, with selection based on
PT suppressing toxic effects of a marker protein.
XX
PS Disclosure; Page 139-140; 201pp; German.
XX
CC The present invention relates to a method for preparing transformed plant
CC cells or organisms, which comprises first transfecting a population of
CC plant cells, already containing at least one marker protein that can
CC exert a direct or indirect toxic effect, with at least one nucleic acid
CC sequence together with at least one compound able to reduce expression,
CC amount, activity and/or function of the marker protein. Transfected cells
CC that contain the nucleic acid in their genome and, because of the
CC activity of the compound have a growth advantage relative to non-
CC transformed cells are selected, under conditions where the marker protein
CC exerts its toxic effect on non-transformed cells. The method is
CC especially used to produce transgenic plants, especially those having an

CC advantageous phenotype, e.g. better nutritional quality for humans or
CC animals, production of selected chemicals or pharmaceuticals, increased
CC resistance to pathogens or environmental stress or higher yields. The
CC present sequence is a coding sequence shown in the exemplification of the
CC invention.
XX
SQ Sequence 720 BP; 172 A; 190 C; 187 G; 171 T; 0 U; 0 Other;
Alignment Scores:
Pred. No.: 2,45e-139 Length: 720
Score: 1222.00 Matches: 238
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 99.58% Mismatches: 0
Query Match: 99.76% Indels: 0
DB: 12 Gaps: 0
US-10-035-300A-2 (1-239) x ADL27841 (1-720)
Qy 1 MetAlaThrProHisIleAsnAlaGluMetGlyAspPheAlaAspValValLeuMetPro 20
Db 1 ATGGCTACCCACACATTAATGCAGAAATGGCGAATTCGTCGACGTAGTTTTTGATGCCA 60
Qy 21 GlyAspProLeuArgAlaLysTyrIleAlaGluThrPheLeuGluAspAlaArgGluVal 40
Db 61 GCGGACCGCTGCGTGCAGATATATTGCTGAAACTTTCTTGAAGATGCCCGTGAAGTG 120
Qy 41 AsnAsnValArgGlyMetLeuGlyPheThrGlyThrTyrLysGlyValArgLysIleSerVal 60
Db 121 AACACGTTGCGCGGTATGCTGGGCTTCACCGGTACTTACAAAGCGCCCAAAATTTCCGTA 180
Qy 61 MetGlyHisGlyValGlyIleProSerCysSerIleTyrThrLysGluLeuIleThrAsp 80
Db 181 ATGGGTACCGGTATGGGTATCCGCTCCTGCTCCTACACCAAGAACTGATCACCAGT 240
Qy 81 PheGlyValLysLysIleIleArgValGlySerCysGlyAlaValLeuProHisValLys 100
Db 241 TTCGGCGTGAAGAAAAATTTATCCGCGTGGGTTCCTGTGGCGCAGTTCTTCCGCGCACGTAAA 300
Qy 101 LeuArgAspValValIleGlyMetGlyAlaCysThrAspSerLysValAsnArgIleArg 120
Db 301 CTGCGGACGCTGTTATCGGTATGGGTGCTGCGGATTCACCGATTCACCAAGTTAACCGATCCGT 360
Qy 121 PheLysAspHisAspPheAlaAlaIleAlaAspPheAspMetValArgAsnAlaValAsp 140
Db 361 TTTAAAGACCATGACTTTGGCGCTATCGCTGACTTCGACATGTTGCGTAACGCGATAGAT 420
Qy 141 AlaAlaLysAlaLeuGlyIleAspAlaArgValGlyAsnLeuPheSerAlaAspLeuPhe 160
Db 421 GCAGCTAAAGCATCTGGGTATGTATGCTCGCGTGGGTAAACCTGTTCTCCGCTGACCTGTTTC 480
Qy 161 TyrSerProAspGlyGluMetPheAspValMetGluLysTyrGlyIleLeuGlyValGlu 180
Db 481 TACTCTCCGACGCGGAAATGTTTCGACGTGATGGAATAATACGGCATTTCTCGCGGTGAA 540
Qy 181 MetGluAlaAlaGlyIleTyrGlyValAlaAlaGluPheGlyAlaLysAlaLeuThrIle 200
Db 541 ATGGAACGGCTGGTATCTACCGCGTGCCTGCAGAAATTTGGCGGAAAGCCCTGACCATC 600
Qy 201 CysThrValSerAspHisIleArgThrHisGluGlnThrThrAlaAlaGluArgGlnThr 220
Db 601 TGCACCGTATCTGACCATCTCCGACTCAGAGCGAGACCACTGCGCGTGAAGCTCAGACT 660
Qy 221 ThrPheAsnAspMetIleLysIleAlaLeuGluSerValLeuLeuGlyAspLysGlu 239
Db 661 ACCTTCAACGACATGATCAAAATCGCACTGGAATCCGTTCTGCTGGCGGATAAAGAG 717
RESULT 5
ABL51595
ID ABL51595 standard; DNA; 752 BP.
XX
AC ABL51595;
XX
DT 04-JUL-2002 (first entry)

XX Purine nucleoside phosphorylase related DNA sequence.
DE
XX Purine nucleoside phosphorylase; enzyme; molecular biology;
KW
XX biotechnology; genetic engineering; gene; ds.
XX
OS Escherichia coli.
OS
XX Synthetic.

XX Key Location/Qualifiers
FH primer_bind 7..35
FT /tag= a
FT /note= "oligonucleotide binding site"
FT complement(709..742)
FT /tag= b
FT /note= "oligonucleotide binding site"

XX RU2179188-C2.

XX 10-FEB-2002.

XX 03-MAR-2000; 2000RU-00105214.

XX 03-MAR-2000; 2000RU-00105214.

XX (ASBI=) AS RUSSIA BIO-ORGANIC CHEM INST.

XX Esipov RS, Gurevich AI, Miroshnikov AI, Chuvikovskii DV;

XX WPI; 2002-265551/31.

XX Production of recombinant purine nucleoside phosphorylase, recombinant
PT plasmid DNA pERPUPH01 and strain Escherichia coli BL21(DE3)perpuh01 for
PT its realization.

XX Disclosure; Col 9-10; Opp; Russian.

XX The present invention describes a recombinant plasmid DNA pERPUPH01 which
CC encodes an amino acid sequence of an Escherichia coli purine nucleoside
CC phosphorylase. The recombinant plasmid DNA pERPUPH01 consists of: a
CC NcoI/EcoRI-fragment of plasmid pET23d DNA containing a promoter and a
CC transcription terminator of T7 RNA polymerase, a translation enhancer of
CC gene 10 of phage T7, a beta-lactamase gene and a NcoI/EcoRI-fragment of a
CC DNA containing the Escherichia coli purine nucleoside phosphorylase gene
CC sequence adapted to these sites. The recombinant plasmid DNA pERPUPH01
CC can be used in molecular biology, biotechnology and genetic engineering.
CC The present sequence represents a nucleotide sequence given in the
CC exemplification of the present invention

XX SQ Sequence 752 BP; 184 A; 194 C; 195 G; 179 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 2,6e-139 Length: 752
Score: 1222.00 Matches: 238
Percent Similarity: 100.00% Conservativeness: 1
Best Local Similarity: 99.58% Mismatches: 0
Query Match: 99.76% Indels: 0
DB: 6 Gaps: 0

US-10-035-300A-2 (1-239) x ABL51595 (1-752)

Qy 1 MetAlaThrProHisIleAsnAlaGluMetGlyAspPheAlaAspValValLeuMetPro 20
Db 13 ATGGCTACCCACACATTAATGCAGAAATGGCGATTTCGCTGAGTAGTTTGTATGCCA 72
Qy 21 GlyAspProLeuArgAlaLysTyrIleAlaGluThrPheLeuGluAspAlaArgGluVal 40
Db 73 GGCGACCCGCTGCGTGGCGAAGTATATTCGCTGAAACTTTCTTGAAGATGCCCGTGAAGTG 132
Qy 41 AsnAsnValArgGlyMetLeuGlyPheThrGlyThrTyrLysGlyArgLysIleSerVal 60
Db 133 AACACAGTTCGCGGTATCTGCGCTTACCCGTTACTTACAAAGGCCGCAAAATTCGTA 192

Qy 61 MetGlyHisGlyValGlyIleProSerCysSerIleTyrThrLysGluLeuIleThrAsp 80
Db 193 ATGGGTACCGTATATGGGTATCCGTCCTGCTCCATCTACACCAAGAACTGATCACCGAT 252
Qy 81 PheGlyValLysValIleIleArgValGlySerCysGlyAlaValLeuProHisValLys 100
Db 253 TTCGCGTGAAGAAATATATCCGCTGGTTCCTGTGGCGCAGTTCGCGGACGTAATAA 312
Qy 101 LeuArgAspValValIleGlyMetGlyAlaCysThrAspSerLysValAsnArgIleArg 120
Db 313 CTGCGCGAGCTCGTTATCGGTATGGTGCCTGCACCGATTCCAAAGTTTAAACCGCATCGT 372
Qy 121 PheLysAspHisAspPheAlaAlaIleAlaAspPheAspMetValArgAsnAlaValAsp 140
Db 373 TTTAAAGACCATGACTTTTGGCGCTATCGCTGACTTCGACATGGTGGCGTAACGCAGTAGAT 432
Qy 141 AlaAlaLysAlaLeuGlyIleAspAlaArgValGlyAsnLeuPheSerAlaAspLeuPhe 160
Db 433 GCAGCTAAAGCAGCTGGGTATTGATGCTGCGGTGGGTAACTGTCTCCGCTGACCTGTTTC 492
Qy 161 TyrSerProAspGlyGluMetPheAspValMetGluLysTyrGlyIleLeuGlyValGlu 180
Db 493 TACTCTCGGACGCGGAAATGTTTCGACGTGATGCAAAATACGGCATTTCTCGGCGTGAA 552
Qy 181 MetGluAlaAlaGlyIleTyrGlyValAlaAlaGluPheGlyAlaLysAlaLeuThrIle 200
Db 553 ATGGAAGCGCTGTTATCTACGGCTGCTGCAGAAATTTGGCGGAAAGCCCTGACCATC 612
Qy 201 CysThrValSerAspHisIleArgThrHisGluGlnThrThrAlaAlaGluArgGlnThr 220
Db 613 TGCACCGTATCTGACCACTCCGCACTACAGCAGACCACTGCGCGTGGCGGTGAGACT 672
Qy 221 ThrPheAsnAspMetIleLysIleAlaLeuGluSerValLeuLeuGlyAspLysGlu 239
Db 673 ACCTTCAACGACATGATCAAAATCGCACTGGAATCCGTTCTGCTGGCGGATAAAGAG 729

RESULT 6

AAAA51632

ID AAA51632 standard; DNA; 3383 BP.

AC AAA51632;

XX 31-OCT-2000 (first entry)

Plasmid pGM678 containing E. coli deoD gene.

XX Uridine phosphorylase; udp; purine nucleoside phosphorylase; PNP; deoD;
KW Genetically modified organism; catalyst; transglycosylation;
KW nucleoside analogue; alpha-pentose-1-phosphate sugar; phosphorolysis;
KW anti-viral; anti-tumour; mesophilic bacterium; ss.

XX Escherichia coli.

OS Synthetic.

XX Key Location/Qualifiers

FH misc_RNA 1..230

FT /tag= a

FT /label= pUC18_sequence

FT CDS 216..952

FT /*tag= b

FT /product= "lacZ-deoD_fused_gene"

FT misc_RNA 961..3383

FT /*tag= c

FT /label= pUC18_sequence

XX WO2000039307-A2.

XX 06-JUL-2000.

XX 23-DEC-1999; 99WO-EP010416.

XX 23-DEC-1998; 98IT-MI002792.

XX

PA (NORP-) NORPHARMA SPA.
XX Bestetti G, Cali' S, Ghisotti D, Orsini G, Tonon G, Zuffi G;
XX WPI; 2000-452402/39.
XX
XX Recombinant expression vectors used to transform cells for the production
PT of e.g. nucleosides encode uridine phosphorylase and/or purine nucleoside
PT phosphorylase.
XX
XX Claim 14; Page 47-48; 72pp; English.
XX
XX Host cells genetically modified to express uridine phosphorylase (Udp)
CC and purine nucleoside phosphorylase (PNP) or their corresponding crude or
CC purified extracts, either separately or in combination are used as
CC catalysts of transglycosylation reactions between a donor nucleoside and
CC an acceptor base, for preparing nucleoside analogues containing
CC heterocyclic systems with purine and/or pyrimidine bases substituted by
CC one or more nitrogen atoms, for preparing alpha-pentose-1-phosphate
CC sugars by phosphorylation reactions and for producing nucleosides and
CC modified nucleoside analogues (all claimed). The modified or natural
CC nucleosides are used directly or as intermediates in the preparation of
CC drugs with anti-viral and anti-tumour activity and for preparing
CC oligonucleotides for therapeutic or diagnostic use. The recombinant host
CC cells are produced using recombinant plasmid expression vectors
CC comprising at least one gene sequence of a mesophilic bacterium coding
CC for a polypeptide with UDP activity and/or enzyme PNP activity and at
CC least one gene sequence coding for antibiotic resistance. In particular,
CC the E. coli udp and deod genes are used. Recombinant strains produced
CC using the vectors express polypeptides with enzyme UDP and PNP activity
CC in large amounts, e.g. 340-1040 times higher UDP activity and 120-200
CC times higher PNP activity than non-transformed corresponding wild type
CC strains
XX
XX SQ Sequence 3383 BP; 826 A; 868 C; 855 G; 834 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 2,07e-138 Length: 3383
Score: 1222.00 Matches: 238
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 99.58% Mismatches: 0
Query Match: 99.76% Indels: 0
DB: 3 Gaps: 0

US-10-035-300A-2 (1-239) x AAAS1632 (1-3383)

Qy 1 MetAlaThrProHisIleAenAlaGluMetGlyAspPheAlaAspValValLeuMetPro 20
Db 240 ATGGCTACCCACACATTAATGCAGAAATGGCGATTTCGCTGACGCTAGTTTGTATGCCA 299
Qy 21 GlyAspProLeuArgAlaIleValThrGlyPheAlaGluThrPheLeuGluAspAlaArgGluVal 40
Db 300 GGGGACCCGCTGGGTGGGAAGTATATGCTGGAACATTTCTTGAAGATGCCCGTGAAGTG 359
Qy 41 AsnAsnValArgGlyMetLeuGlyPheThrGlyThrTyrlsGlyArgIleSerVal 60
Db 360 AACACGTTCCGGGTATGCTGGGCTTCACCGTACTTACAAAGCCGCAAAATTTCCGTA 419
Qy 61 MetGlyHisGlyValGlyIleProSerCysSerIleTyrlsGlyGluLeuIleThrAsp 80
Db 420 ATGGGTACCGGTATGGGTATCCGCTCCGCTCCATCTACACCAAGAACATGATCACCAGT 479
Qy 81 PheGlyValIleValIleIleArgValGlySerCysGlyAlaValLeuProHisValIle 100
Db 480 TTCGGGTGAAGAAATATATCCGGTGGGTTCCTGTGGCGCAGTTCTGGCGCACGTAAAA 539
Qy 101 LeuArgAspValValIleGlyMetGlyValAlaCysThrAspSerIleValAsnArgIleArg 120
Db 540 CTGGCGACGCTCGTTATCGGTATGGGTGCTTCACCGGATTCGAAGTTAACCGCATCCGT 599
Qy 121 PheIleAspHisAspPheAlaIleAlaAspPheAspMetValArgAsnAlaValAsp 140
Db 600 TTTTAAAGACCATGACTTTGCGGTATCGCTGACTTCGACATGTTGCGTAACGCGTAGAT 659

Qy 141 AlaAlaIleValLeuGlyIleAspAlaArgValGlyAsnLeuPheSerAlaAspLeuPhe 160
Db 660 GCAGCTAAAGCAGCTGGGTATGATGCTCGCGTGGGAACCTGTTCTCCGCTACCTGTTTC 719
Qy 161 TyrSerProAspGlyGluMetPheAspValMetGluIleTyrlsGlyIleLeuGlyValGlu 180
Db 720 TACTCTCCGACGCGCAAAATGTTTCGACGCTGCTGCAGAAATTTGGCGGAAAGCCCTGACC 779
Qy 181 MetGluAlaAlaGlyIleTyrlsGlyValAlaAlaGluPheGlyAlaIleValLeuThrIle 200
Db 780 ATGGAAGCGCTGCTATCTACGCGCTGCTGCAGAAATTTGGCGGAAAGCCCTGACCATC 839
Qy 201 CysThrValSerAspHisIleArgThrHisGluGlnThrThrAlaAlaGluArgGlnThr 220
Db 840 TGCACCGTATCTGACCAATCCGCTCAGACGACACCACTCCCGCTGAGCGTACAGACT 899
Qy 221 ThrPheAsnAspMetIleIleAlaLeuGluSerValLeuLeuGlyAspIleGlu 239
Db 900 ACCTTCAACGACATGATCAAAATCGCACTTGGATCGTCTGCTGGCGGATAAAGAG 956
RESULT 7
AAAS1634
ID AAAS1634 standard; DNA; 4189 BP.
XX
AC AAAS1634;
XX 31-OCT-2000 (first entry)
XX Plasmid pGM712 containing udp and deod genes.
DE Uridine phosphorylase; udp; purine nucleoside phosphorylase; PNP; deod;
KW Genetically modified organism; catalyst; transglycosylation; anti-viral;
KW nucleoside analogue; alpha-pentose-1-phosphate sugar; phosphorylase;
KW anti-tumour; mesophilic bacterium; tetracycline resistance; ss.
XX Escherichia coli.
OS Synthetic.
OS Chimeric.
XX
FH Key Location/Qualifiers
FT misc_RNA 1. .242
FT /tag= a
FT /label= pUC18_sequence
FT 243. .1021
FT /tag= b
FT /label= udp_gene_sequence
FT 1022. .1025
FT /tag= c
FT /label= pUC18_sequence
FT 1026. .1036
FT /tag= d
FT /label= pBAD24_sequence
FT 1037. .1766
FT /tag= e
FT /label= deod_gene_sequence
FT 1767. .1792
FT /tag= f
FT /label= pBAD24_sequence
FT 1793. .4189
FT /tag= g
FT /label= pUC18_sequence
XX
XX WO200039307-A2.
XX
XX 06-JUL-2000.
XX
XX 23-DEC-1999; 99WO-EP010416.
XX
XX 23-DEC-1998; 98IT-MI002792.
XX (NORP-) NORPHARMA SPA.
XX

PI Bestetti G, Cali' S, Ghisotti D, Orsini G, Tonon G, Zuffi G;
 XX WPI; 2000-452402/39.
 XX Recombinant expression vectors used to transform cells for the production
 PT of e.g. nucleosides encode uridine phosphorylase and/or purine nucleoside
 PT phosphorylase.
 XX
 XX Claim 14; Page 51-52; 72pp; English.
 XX
 CC Host cells genetically modified to express uridine phosphorylase (Udp)
 CC and purine nucleoside phosphorylase (PNP) or their corresponding crude or
 CC purified extracts, either separately or in combination are used as
 CC catalysts of transglycosylation reactions between a donor nucleoside and
 CC an acceptor base, for preparing nucleoside analogues containing
 CC heterocyclic systems with purine and/or pyrimidine bases substituted by
 CC one or more nitrogen atoms, for preparing alpha-pentose-1-phosphate
 CC sugars by phosphorolysis reactions and for producing nucleosides and
 CC modified nucleoside analogues (all claimed). The modified or natural
 CC nucleosides are used directly or as intermediates in the preparation of
 CC drugs with anti-viral and anti-tumour activity and for preparing
 CC oligonucleotides for therapeutic or diagnostic use. The recombinant host
 CC cells are produced using recombinant plasmid expression vectors
 CC comprising at least one gene sequence of a mesophilic bacterium coding
 CC for a polypeptide with UDP activity and/or enzyme PNP activity and at
 CC least one gene sequence coding for antibiotic resistance. In particular,
 CC the E. coli udp and deo genes are used. Recombinant strains produced
 CC using the vectors express polypeptides with enzyme UDP and PNP activity
 CC in large amounts, e.g. 340-1040 times higher UDP activity and 120-200
 CC times higher PNP activity than non-transformed corresponding wild type
 CC strains
 XX
 XX Sequence 4189 BP; 998 A; 1077 C; 1081 G; 1033 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 2,796-138 Length: 4189
 Score: 1222.00 Matches: 238
 Percent Similarity: 100.00% Conservative: 1
 Best Local Similarity: 99.58% Mismatches: 0
 Query Match: 99.76% Indels: 0
 DB: 3 Gaps: 0

US-10-035-300A-2 (1-239) x AAAS1634 (1-4189)

Qy 1 MetAlaThrProHisIleAsnAlaGluMetGlyAspPheAlaAspValValLeuMetPro 20
 Db 1046 ATGGCTACCCACACATTAATGCAGAAATGGCGGATTTTCGTCAGCTAGTTTTCATGCCA 1105

Qy 21 GlyAspProLeuArgAlaLysTyrIleAlaGluThrPheLeuGluAspAlaArgGluVal 40
 Db 1106 GCGGACCGCTGCGTGGAGATATATTCGTGAACCTTTTCCTGAAGATGCCCGTGAAGTG 1165

Qy 41 AsnAsnValArgGlyMetLeuGlyPheThrGlyThrTyrLysGlyValGlySerVal 60
 Db 1166 AACACAGTTCGCGGTATCTGGCTTACCGGTACTTACAAGGCCGCAAAATTTCCGTA 1225

Qy 61 MetGlyHisGlyValGlyIleProSerCysSerIleTyrThrLysGluLeuIleThrAsp 80
 Db 1226 ATGGGTACCGTATGGGTATCCCGTCTCTGCTCCATCTACACCAAGAACTGATCACCAGT 1285

Qy 81 PheGlyValLysLysIleIleArgValGlySerCysGlyAlaValLeuProHisValLys 100
 Db 1286 TTCGCGGTGAAGAAATATATCCGCTGGGTTCCTGTGGCGAGTTCTGCCGACGTAAA 1345

Qy 101 LeuArgAspValIleGlyMetGlyAlaCysThrAspSerLysValAsnArgIleArg 120
 Db 1346 CTGGCGACGTCGTATTCGTTATGGGTGCGTCCGACCGGATTCAAAGTTAAACCGATCCGT 1405

Qy 121 PheLysAspHisAspPheAlaAlaIleAlaAspPheAspMetValArgAsnAlaValAsp 140
 Db 1406 TTTAAAGACCATGACTTTTCGCGCTATCGCTGACTTCGACATGGTGGCTAACGAGTAGAT 1465

Qy 141 AlaAlaLysAlaLeuGlyIleAspAlaArgValGlyAsnLeuPheSerAlaAspLeuPhe 160

Db 1466 GCACCTAAAGCAGCTGGGTATTGATGCTCGCGGTGGGTAACTTCTCCGCTGACCTGTC 1525
 Qy 161 TyrSerProAspGlyGluMetPheAspValMetGluLysTyrGlyIleLeuGlyValGlu 180
 Db 1526 TACTCTCGGACGGCGAAATGTTTCAGCGTGTATGAAAAAATACGGCATTTCTGGCGTGAA 1585
 Qy 181 MetGluAlaAlaGlyIleTyrGlyValAlaAlaGluPheGlyAlaLysAlaLeuThrIle 200
 Db 1586 ATGGAAGCGGCTGGTATCTACGGGTCGCTGACAGATTTGGCGGAAAGCCCTGACCATC 1645
 Qy 201 CysThrValSerAspHisIleArgThrHisGluGlnThrThrAlaAlaGluArgGlnThr 220
 Db 1646 TGCACCGTATCTGACCAATCCGCACTCACGAGCAGACCACTGCGCGTGCAGCGTCAGACT 1705
 Qy 221 ThrPheAsnAspMetIleLeuAlaLeuGluSerValLeuLeuGlyAspLysGlu 239
 Db 1706 ACCTTCAACGACATGATCAAAATCGCACTGGATCCGTTCTGCTGGGCGATAAAGAG 1762

RESULT 8
 ACA63355
 ID ACA63355 standard; DNA; 5013 BP.
 XX ACA63355;
 AC ACA63355;
 XX 29-AUG-2003 (first entry)
 XX E. coli DNA encoding PNP inserted in plasmid pTRCNP.
 XX Prokaryotic host cell; transformation; transfection; vector; PNP;
 XX purine nucleoside phosphorylase; hydrolase; replicating cell;
 XX nonreplicating cell; bystander cell; purine substrate; tumour cell;
 XX virally infected cell; cancer therapy; cytostatic; mutant; pTRCNP; ds.
 XX Escherichia coli; strain JM101.
 OS Synthetic.
 XX US6491905-B1.
 XX 10-DEC-2002.
 XX 30-OCT-1998; 98US-00183188.
 XX 14-SEP-1993; 93US-00123231.
 PR 23-AUG-1996; 96US-00702181.
 PR 24-JUN-1997; 97US-00881772.
 PR 31-OCT-1997; 97US-0064676P.
 XX (UABR-) UAB RES FOUND.
 PA (SOUR) SOUTHERN RES INST.
 PI Sorscher EJ, Parker WB, Waud W, Gadi VK, Bennett LL;
 XX WPI; 2003-502635/47.
 XX Novel prokaryotic host cell for treating tumor cells and virally infected
 PT cells transfected by a vector comprising a DNA sequence encoding Escherichia
 PT encoding Escherichia coli derived purine nucleoside phosphorylase.
 XX Claim 1; Fig 15; 45pp; English.
 XX The present invention relates to a prokaryotic host cell transformed or
 CC transfected by a vector comprising a DNA sequence encoding Escherichia
 CC coli derived purine nucleoside phosphorylase (PNP) or hydrolase. The
 CC vector is useful for killing targeted replicating or nonreplicating
 CC mammalian cells, and bystander cells, the host cell is useful in
 CC combination with a purine substrate for treating tumour and/or virally
 CC infected cells. The vector is useful in cancer therapy. The present
 CC sequence represents DNA encoding E.coli PNP inserted in plasmid pTRCNP
 SQ Sequence 5013 BP; 1231 A; 1258 C; 1325 G; 1199 T; 0 U; 0 Other;

Alignment Scores:

```
Pred. No.: 3.57e-138 Length: 5013
Score: 1222.00 Matches: 238
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 99.58% Mismatches: 0
Query Match: 99.76% Indels: 0
DB: 9 Gaps: 0

US-10-035-300A-2 (1-239) x ACA63355 (1-5013)

Qy 1 MetAlaThrProHisIleAsnAlaGluMetGlyAspPheAlaAspValValLeuMetPro 20
Db 413 ATGGCTACCCACACATTAATGCGAATATGGCGGATTCGCTGACGTAGTTTGTATGCCA 472

Qy 21 GlyAspProLeuArgAlaIleValThrIleAlaGluThrPheLeuGluAspAlaArgGluVal 40
Db 473 GCGACCGCTGCGTGGAGATATATGCTGAACATTTCTTGAAGATGCCGGAAGTG 532

Qy 41 AsnAsnValArgGlyMetLeuGlyPheThrGlyThrTyrlsGlyArgLysIleSerVal 60
Db 533 AACACGTTCCGGGTATGCTGGGCTTCACCGGTACTTACAAAGCGCGCAAAATTTCCGTA 592

Qy 61 MetGlyHisGlyValGlyIleProSerCysSerIleTyrlsGlyLeuLeuIleThrAsp 80
Db 593 ATGGGTACCGGTATGGGTATCCCGTCTCTCCATCTACACCAAGAACTGATCACCAGT 652

Qy 81 PheGlyValLysIleIleArgValGlySerCysGlyAlaValLeuProHisValLys 100
Db 653 TTCGGCGTGAAGAAATATTCGGCTGGGTTCCTGTGGCGGAGTTCCTGCCGACGTAAAA 712

Qy 101 LeuArgAspValIleGlyMetGlyAlaCysThrAspSerLysValAsnArgIleArg 120
Db 713 CTGGCGACGTCGTTATCGGTATGGGTGCTGACCGCATTCCAAAGTTAACCGCATCCGT 772

Qy 121 PheLysAspHisAspPheAlaIleAlaAspPheAspMetValArgAsnAlaValAsp 140
Db 773 TTTAAAGACCATGACTTTCGGCTATCGCTGACTTCGACATGGTGGCTAAACGCGATGAT 832

Qy 141 AlaAlaLysAlaLeuGlyIleAspAlaArgValGlyAsnLeuPheSerAlaAspLeuPhe 160
Db 833 GCAGCTAAAGCAGTGGGTATGTATGCTGCGTGGGTAACTGTCTTCGCTGACCTGTTTC 892

Qy 161 TyrSerProAspGlyGluMetPheAspValMetGluLysTyrlsGlyIleLeuGlyValGlu 180
Db 893 TACTCTCCGACGCGGAAATGTTCCAGCGTGATGGAATAACGCAATTCCTCGGCGTGGAA 952

Qy 181 MetGluAlaAlaGlyIleTyrlsGlyValAlaAlaGluPheGlyAlaLysAlaLeuThrIle 200
Db 953 ATGGAACGCGCTGCTATCTACGCGCTCGCTGCAGAAATTTGGCGCGAAAGCCCTGACCATC 1012

Qy 201 CysThrValSerAspHisIleArgThrHisGluGlnThrThrAlaAlaGluArgGlnThr 220
Db 1013 TGCACCGTATCTGACCACTCCGACATCAGACGACGACCACTGCGCGTGAAGCTCAGACT 1072

Qy 221 ThrPheAsnAspMetIleLysIleAlaLeuGluSerValLeuLeuGlyAspLysGlu 239
Db 1073 ACCTTCAACGACATGATCAAAATCGCACTGGAATCCGTTCTGCTGGCGCATAAAGAG 1129

RESULT 9
AAD59423
ID AAD59423 standard; DNA; 5013 BP.
XX
AC
AD59423;
XX
XX
18-DEC-2003 (first entry)
XX
DE E.coli purine analogue nucleoside phosphorylase DNA.
XX
KW Cytostatic; virucide; purine cleavage enzyme; tumour; therapy; PNP;
KW drug screening; purine analogue nucleoside phosphorylase; ds.
XX
OS Escherichia coli.
XX
PN US2003077268-A1.
XX
```

```
XX 24-APR-2003.
PD 18-JUL-2002; 2002US-00198034.
XX
PF 14-SEP-1993; 93US-00122321.
XX 23-AUG-1996; 96US-00702181.
PR 24-JUN-1997; 97US-00881772.
PR 31-OCT-1997; 97US-0064676P.
PR 30-OCT-1998; 98US-00183188.
XX
XX (SORS/) SORSCHER E J.
PA (PARK/) PARKER W B.
PA (WAUD/) WAUD W.
PA (GADI/) GADI V K.
PA (BENN/) BENNETT L L.
XX
XX Sorscher EJ, Parker WB, Waud W, Gadi VK, Bennett LL;
XX WPI; 2003-755064/71.
XX
XX Killing (non-)replicating targeted mammalian cells and bystander cells
PT useful for treating tumor cells and/or virally infected cells comprises
PT delivering purine nucleotide phosphorylase or hydrolase cleavage enzyme
PT to cells.
XX
XX Disclosure; Page 20-23; 40pp; English.
XX
XX The present invention relates to a method of killing replicating or non-
CC replicating, targeted mammalian cells and bystander cells, comprising
CC delivering a purine cleavage enzyme to the targeted mammalian cells and
CC contacting the targeted cells with a purine cleavage enzyme substrate to
CC kill the targeted cells when cleaved by the enzyme. The purine cleavage
CC enzyme is used for killing replicating or non-replicating targeted
CC mammalian cells and bystander cells. They are used for treating tumour
CC cells and virally infected cells. The method may also be useful in drug
CC screening. This sequence represents Escherichia coli PNP (purine analogue
CC nucleoside phosphorylase) DNA
XX
XX Sequence 5013 BP; 1231 A; 1258 C; 1325 G; 1199 T; 0 U; 0 Other;
SQ
```

```
Alignment Scores:
Pred. No.: 3.57e-138 Length: 5013
Score: 1222.00 Matches: 238
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 99.58% Mismatches: 0
Query Match: 99.76% Indels: 0
DB: 10 Gaps: 0

US-10-035-300A-2 (1-239) x AAD59423 (1-5013)

Qy 1 MetAlaThrProHisIleAsnAlaGluMetGlyAspPheAlaAspValValLeuMetPro 20
Db 413 ATGGCTACCCACACATTAATGCGAATATGGCGGATTCGCTGACGTAGTTTGTATGCCA 472

Qy 21 GlyAspProLeuArgAlaIleValThrIleAlaGluThrPheLeuGluAspAlaArgGluVal 40
Db 473 GCGACCGCTGCGTGGAGATATATGCTGAACATTTCTTGAAGATGCCGGAAGTG 532

Qy 41 AsnAsnValArgGlyMetLeuGlyPheThrGlyThrTyrlsGlyArgLysIleSerVal 60
Db 533 AACACGTTCCGGGTATGCTGGGCTTCACCGGTACTTACAAAGCGCGCAAAATTTCCGTA 592

Qy 61 MetGlyHisGlyValGlyIleProSerCysSerIleTyrlsGlyLeuLeuIleThrAsp 80
Db 593 ATGGGTACCGGTATGGGTATCCCGTCTCTCCATCTACACCAAGAACTGATCACCAGT 652

Qy 81 PheGlyValLysIleIleArgValGlySerCysGlyAlaValLeuProHisValLys 100
Db 653 TTCGGCGTGAAGAAATATTCGGCTGGGTTCCTGTGGCGGAGTTCCTGCCGACGTAAAA 712

Qy 101 LeuArgAspValIleGlyMetGlyAlaCysThrAspSerLysValAsnArgIleArg 120
Db 713 CTGGCGACGTCGTTATCGGTATGGGTGCTGACCGCATTCCAAAGTTAACCGCATCCGT 772

Qy 121 PheLysAspHisAspPheAlaIleAlaAspPheAspMetValArgAsnAlaValAsp 140
Db 773 TTTAAAGACCATGACTTTCGGCTATCGCTGACTTCGACATGGTGGCTAAACGCGATGAT 832

Qy 141 AlaAlaLysAlaLeuGlyIleAspAlaArgValGlyAsnLeuPheSerAlaAspLeuPhe 160
Db 833 GCAGCTAAAGCAGTGGGTATGTATGCTGCGTGGGTAACTGTCTTCGCTGACCTGTTTC 892

Qy 161 TyrSerProAspGlyGluMetPheAspValMetGluLysTyrlsGlyIleLeuGlyValGlu 180
Db 893 TACTCTCCGACGCGGAAATGTTCCAGCGTGATGGAATAACGCAATTCCTCGGCGTGGAA 952

Qy 181 MetGluAlaAlaGlyIleTyrlsGlyValAlaAlaGluPheGlyAlaLysAlaLeuThrIle 200
Db 953 ATGGAACGCGCTGCTATCTACGCGCTCGCTGCAGAAATTTGGCGCGAAAGCCCTGACCATC 1012

Qy 201 CysThrValSerAspHisIleArgThrHisGluGlnThrThrAlaAlaGluArgGlnThr 220
Db 1013 TGCACCGTATCTGACCACTCCGACATCAGACGACGACCACTGCGCGTGAAGCTCAGACT 1072

Qy 221 ThrPheAsnAspMetIleLysIleAlaLeuGluSerValLeuLeuGlyAspLysGlu 239
Db 1073 ACCTTCAACGACATGATCAAAATCGCACTGGAATCCGTTCTGCTGGCGCATAAAGAG 1129

RESULT 9
AAD59423
ID AAD59423 standard; DNA; 5013 BP.
XX
AC
AD59423;
XX
XX
18-DEC-2003 (first entry)
XX
DE E.coli purine analogue nucleoside phosphorylase DNA.
XX
KW Cytostatic; virucide; purine cleavage enzyme; tumour; therapy; PNP;
KW drug screening; purine analogue nucleoside phosphorylase; ds.
XX
OS Escherichia coli.
XX
PN US2003077268-A1.
XX
```

```
Db 713 CTGGCGGACGTCGTTATTCGGTATCGGTGCTGCACCGGATTCCAAAGTTAACCGCATCCGT 772
Qy 121 PheLysAspHisAspPheAlaAlaIleAlaAspPheAspMetValArgAsnAlaValAsp 140
Db 773 TTTAAAGACCATGACTTTGGCGCTATCGCTGACATGGTGGTAACGAGTAGAT 832
Qy 141 AlaAlaLysAlaLeuGlyIleAspAlaArgValGlyAsnLeuPheSerAlaAspLeuPhe 160
Db 833 GCAGCTAAAGCACTGGGTATTGATGCTCGCGTGGGTAACTGTTCTCGCGTACCGTGTTC 892
Qy 161 TyrSerProAspGlyGluMetPheAspValMetGluLysTyrGlyIleLeuGlyValGlu 180
Db 893 TACTCTCCGGACGGCGAAATGTTCCAGCTGATGGAATAATACGGCAATTCCTGGCGTGGAA 952
Qy 181 MetGluAlaAlaGlyIleTyrGlyValAlaAlaGluPheGlyAlaLysAlaLeuThrIle 200
Db 953 ATGGAGCGGCTGGTATCTACGGCGTCTGCAGAAATTTGGCGGAAGCCCTGACCATC 1012
Qy 201 CysThrValSerAspHisIleArgThrHisGluGlnThrThrAlaAlaGluArgGlnThr 220
Db 1013 TGCACCGTATCTGACCATCGCACTCAGCAGCAGACCACTGCCGCTGAGCGTCAGACT 1072
Qy 221 ThrPheAsnAspMetIleLysIleAlaLeuGluSerValLeuGlyAspLysGlu 239
Db 1073 ACCTTCAACGACATGATCAAATTCGACTGGAAATCCGTTCTGCTGGCGGATAAAGAG 1129

RESULT 10
ADE86127
ID ADE86127 standard; DNA; 5013 BP.
XX
AC ADE86127;
XX
DT 29-JAN-2004 (first entry)
XX
DE E. coli purine nucleoside phosphorylase (PNP) gene.
XX
KW Tumour growth inhibition; purine analogue; stable transformation;
KW prokaryotic host cell; purine nucleoside phosphorylase; PNP;
KW purine nucleoside hydrolase; tumour cell; virally infected cell;
KW toxin gene therapy; cytostatic; virucide; gene; ds.
XX
OS Escherichia coli.
XX
FN US2003134819-A1.
XX
PD 17-JUL-2003.
XX
PF 18-JUL-2002; 2002US-00198228.
XX
PR 14-SEP-1993; 93US-00122321.
PR 23-AUG-1996; 96US-00702181.
PR 24-JUN-1997; 97US-00881772.
PR 31-OCT-1997; 97US-0064676P.
PR 30-OCT-1998; 98US-00183186.
XX
PA (SORS/) SORSCHER E J.
PA (PARK/) PARKER W B.
PA (WAUD/) WAUD W.
PA (GADI/) GADI V K.
PA (BENN/) BENNETT L L.
XX
XX Sorscher EJ, Parker WB, Waud W, Gadi VK, Bennett LL;
XX WPI; 2003-897265/82.
XX
XX Use of purine analogs for inhibiting tumor growth and treating virally
XX infected cells.
XX
XX Example 25; Page 20-23; 40pp; English.
XX
CC The present invention relates to a method of inhibiting tumour growth by
CC administration of a purine analogue. The method comprises the stable
CC transformation of a prokaryotic host cell with a vector including a DNA
```

```
CC sequence for purine nucleoside phosphorylase (PNP) or hydrolase. The
CC transformed prokaryotic host cell is used in combination with a purine
CC substrate to treat tumour and/or virally infected cells. The method
CC provides toxin gene therapy to overcome the problem of inefficient
CC delivery, cell replication dependent killing and low toxin diffusion
CC between cells. The present sequence represents the Escherichia coli
CC purine nucleoside phosphorylase (PNP) gene.
XX
SQ Sequence 5013 BP; 1231 A; 1258 C; 1325 G; 1199 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 3.57e-138 Length: 5013
Score: 1222.00 Matches: 238
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 99.58% Mismatches: 0
Query Match: 99.76% Indels: 0
DB: 10 Gaps: 0

US-10-035-300A-2 (1-239) x ADE86127 (1-5013)
Qy 1 MetAlaThrProHisIleAsnAlaGluMetGlyAspPheAlaAspValValLeuMetPro 20
Db 413 ATGGCTACCCACACATTAATGCAGAAATGGCGGATTTGCTGACGTAGTTTGTATGCCA 472
Qy 21 GlyAspProLeuArgAlaLysTyrIleAlaGluThrPheLeuGluAspAlaArgGluVal 40
Db 473 GCGACCCGCTGCGTGGCAAGTATATTGCTGAAAATTTTCTTGAAGATGCCCGTGAAGTG 532
Qy 41 AsnAsnValArgGlyMetLeuGlyPheThrGlyThrTyrLysGlyValLysIleSerVal 60
Db 533 AACAACTGTCGCGGTATGCTGGGCTTACCGGATCTTACAAGGCGCGCAAAATTTCCGTA 592
Qy 61 MetGlyHisGlyValGlyIleProSerCysSerIleTyrThrLysGluLeuIleThrAsp 80
Db 593 ATGGGTACCGTATGGGTATCCCGTCTGCTCCATCTACACCAAGAACTGATCACCGAT 652
Qy 81 PheGlyValLysLysIleAlaArgValGlySerCysGlyAlaValLeuProHisValLys 100
Db 653 TTCGCGGTGAAGAAATATATCCGCTGGGTTCTCTGTGGCGAGTTCTCGCGCACGTAANA 712
Qy 101 LeuArgAspValValIleGlyMetGlyAlaCysThrAspSerLysValAsnArgIleArg 120
Db 713 CTGGCGACGCTGGTATCGGTATGGGTGCTGCACCGGATTCCAAAGTTAACCGCATCCGT 772
Qy 121 PheLysAspHisAspPheAlaAlaIleAlaAspPheAspMetValArgAsnAlaValAsp 140
Db 773 TTTAAAGACCATGACTTTGGCGCTATCGCTGACATGGTGGTGAACGACAGTAGAT 832
Qy 141 AlaAlaLysAlaLeuGlyIleAspAlaArgValGlyAsnLeuPheSerAlaAspLeuPhe 160
Db 833 GCAGCTAAAGCACTGGGTATTGATGCTCGCGTGGGTAACTGTTCTCGCGTACCGTGTTC 892
Qy 161 TyrSerProAspGlyGluMetPheAspValMetGluLysTyrGlyIleLeuGlyValGlu 180
Db 893 TACTCTCCGGACGGCGAAATGTTCCAGCTGATGGAATAATACGGCAATTCCTGGCGTGGAA 952
Qy 181 MetGluAlaAlaGlyIleTyrGlyValAlaAlaGluPheGlyAlaLysAlaLeuThrIle 200
Db 953 ATGGAGCGGCTGGTATCTACGGCGTCTGCAGAAATTTGGCGGAAGCCCTGACCATC 1012
Qy 201 CysThrValSerAspHisIleArgThrHisGluGlnThrThrAlaAlaGluArgGlnThr 220
Db 1013 TGCACCGTATCTGACCATCGCACTCAGCAGCAGACCACTGCCGCTGAGCGTCAGACT 1072
Qy 221 ThrPheAsnAspMetIleLysIleAlaLeuGluSerValLeuGlyAspLysGlu 239
Db 1073 ACCTTCAACGACATGATCAAATTCGACTGGAAATCCGTTCTGCTGGCGGATAAAGAG 1129

RESULT 11
AAA51633
ID AAA51633 standard; DNA; 5495 BP.
XX
AC AAA51633;
```

```
XX 31-OCT-2000 (first entry)
XX Plasmid pGM707 containing deoD and tetracycline resistance genes.
XX
XX Uridine phosphorylase; udp; purine nucleoside phosphorylase; PNP; deoB;
XX Genetically modified organism; catalyst; transglycosylation; anti-viral;
XX nucleoside analogue; alpha-pentose-1-phosphate sugar; phosphorolysis;
XX anti-tumour; mesophilic bacterium; tetracycline resistance; ss.
XX
XX Escherichia coli.
XX Synthetic.
XX Chimeric.
XX
XX Key Location/Qualifiers
XX misc_RNA 1..230
XX /tag= a
XX /label= pUC18_sequence
XX CDS 216..952
XX /tag= b
XX /product= "lacZ-udp_fused_gene"
XX misc_RNA 961..978
XX /tag= c
XX /label= pUC18_sequence
XX misc_RNA 979..1422
XX /tag= d
XX /label= pHP45_sequence
XX misc_RNA 1423..2822
XX /tag= e
XX /label= pBR322_Tet_gene_sequence
XX misc_RNA 2823..3090
XX /tag= f
XX /label= pHP45_sequence
XX misc_RNA 3091..5495
XX /tag= g
XX /label= pUC18_sequence
XX
XX WO200039307-A2.
XX
XX 06-JUN-2000.
XX
XX 23-DEC-1999; 99WO-EP010416.
XX
XX 23-DEC-1998; 98IT-MI002792.
XX
XX (NORP-) NORPHARMA SPA.
XX
XX Bestetti G, Cali' S, Ghisotti D, Orsini G, Tonon G, Zuffi G;
XX WPI; 2000-452402/39.
XX
XX Recombinant expression vectors used to transform cells for the production
XX of e.g. nucleosides encode uridine phosphorylase and/or purine nucleoside
XX phosphorylase.
XX
XX Claim 14; Page 48-50; 72pp; English.
XX
XX Host cells genetically modified to express uridine phosphorylase (Udp)
XX and purine nucleoside phosphorylase (PNP) or their corresponding crude or
XX purified extracts, either separately or in combination are used as
XX catalysts of transglycosylation reactions between a donor nucleoside and
XX an acceptor base, for preparing nucleoside analogues containing
XX heterocyclic systems with purine and/or pyrimidine bases substituted by
XX one or more nitrogen atoms, for preparing alpha-pentose-1-phosphate
XX sugars by phosphorolysis reactions and for producing nucleosides and
XX modified nucleoside analogues (all claimed). The modified or natural
XX nucleosides are used directly or as intermediates in the preparation of
XX drugs with anti-viral and anti-tumour activity and for preparing
XX oligonucleotides for therapeutic or diagnostic use. The recombinant host
XX cells are produced using recombinant plasmid expression vectors
XX comprising at least one gene sequence of a mesophilic bacterium coding
XX for a polypeptide with UDP activity and/or enzyme PNP activity and at
XX least one gene sequence coding for antibiotic resistance. In particular,
```

```
CC the E. coli udp and deoD genes are used. Recombinant strains produced
CC using the vectors express polypeptides with enzyme udp and PNP activity
CC in large amounts, e.g. 340-1040 times higher udp activity and 120-200
CC times higher PNP activity than non-transformed corresponding wild type
CC strains
XX
XX SQ Sequence 5495 BP; 1335 A; 1416 C; 1460 G; 1284 T; 0 U; 0 Other;
XX
XX Alignment Scores:
XX Pred. No.: 4.05e-138 Length: 5495
XX Score: 1222.00 Matches: 238
XX Percent Similarity: 100.00% Conservative: 1
XX Best Local Similarity: 99.58% Mismatches: 0
XX Query Match: 99.76% Indels: 0
XX DB: 3 Gaps: 0
XX
XX US-10-035-300A-2 (1-239) x AAAS1633 (1-5495)
XX
XX QY 1 MetAlaThrProHisIleAsnAlaGluMetGlyAspPheAlaAspValValLeuMetPro 20
XX DB 240 ATGCTACCCACACACATTATATGAGAAATGGCGATTTCCTGACGTAGTTTGTATGCCA 299
XX
XX QY 21 GlyAspProLeuArgAlaLysTyrIleAlaGluThrPheLeuGluAspAlaArgGluVal 40
XX DB 300 GGGCACCCGCTGGGTGCGAAGTATATTGCTGAAAACTTTCCTTGAAGATGCCGTGAAGTG 359
XX
XX QY 41 AsnAsnValArgGlyMetLeuGlyPheThrGlyThrTyrLysGlyArgLysIleSerVal 60
XX DB 360 AACACGTTCCGCGTATGCTGGGCTTCACCGGTACTTACAAAGCCGCAAAATTTCCGTA 419
XX
XX QY 61 MetGlyHisGlyValGlyIleProSerCysSerIleTyrThrLysGluLeuIleThrAsp 80
XX DB 420 ATGGTCACGGTATGGGTATCCCGTCTCTCCATCTACACCAAGAACTGATCACCAGT 479
XX
XX QY 81 PheGlyValLysLysIleIleArgValGlySerCysGlyAlaValLeuProHisValLys 100
XX DB 480 TTCGGCGTGAAGAAAATTTATCCGCTGGGTTCCTGTGGCGCAGTTCTGCCGCAAGTAAAA 539
XX
XX QY 101 LeuArgAspValValIleGlyMetGlyAlaCysThrAspSerLysValAsnArgIleArg 120
XX DB 540 CTGGCGACGTCGTATTTCGGTATGGGTGCTGCCGATTCACCGATTCCAAAGTTAACCGCATCCGT 599
XX
XX QY 121 PheLysAspHisAspPheAlaIleAlaAspPheAspMetValArgAsnAlaValAsp 140
XX DB 600 TTTAAAGACCATGACTTTTCCCGCTATCGCTGACTTCGACATGGTGGCTAACCGCAGTAGAT 659
XX
XX QY 141 AlaAlaLysAlaLeuGlyIleAspAlaArgValGlyAsnLeuPheSerAlaAspLeuPhe 160
XX DB 660 GCAGCTAAAGCACTGGGTATTGATGCTCGGTGAGTAACCTGTTCTCCGCTGACCTGTTTC 719
XX
XX QY 161 TyrSerProAspGlyGluMetPheAspValMetGluLysTyrGlyIleLeuGlyValGlu 180
XX DB 720 TATCTCCGACCGCGGAAATGTTTCGACGTGATGGAAAAAATACGGCATTCCTCGGGGTGAA 779
XX
XX QY 181 MetGluAlaAlaGlyIleTyrGlyValAlaAlaGluPheGlyAlaLysAlaLeuThrIle 200
XX DB 780 ATGGAAGCGCTGGTATCTACGGCGTCGCTGCAGATTTCGGCGGAAAGCCCTGACCATC 839
XX
XX QY 201 CysThrValSerAspHisIleArgThrHisGluGlnThrThrAlaAlaGluArgGlnThr 220
XX DB 840 TGCACCGTATCTGACCACTCCGCACTCAGCAGCAGACCACTCCGCTGAGCGTCAGACT 899
XX
XX QY 221 ThrPheAsnAspMetIleLysIleAlaLeuGluSerValLeuLeuGlyAspLysGlu 239
XX DB 900 ACCTTCACGACATGATCAAAATCGCACTGGAAATCCGTTCTGCTGGGCGGATTAAGAG 956
XX
XX RESULT 12
XX AAAS1635
XX ID AAAS1635 standard; DNA; 6301 BP.
XX
XX AC AAAS1635;
XX
XX DT 31-OCT-2000 (first entry)
```

XX DE Plasmid pGM716 containing udp, deoD and tetracycline resistance genes.

XX KW Uridine phosphorylase; udp; purine nucleoside phosphorylase; PNP; deoD;

KW Genetically modified organism; catalyst; transglycosylation; anti-viral;

KW nucleoside analogue; alpha-pentose-1-phosphate sugar; phosphorolysis;

XX anti-tumour; mesophilic bacterium; tetracycline resistance; ss.

XX OS Escherichia coli.

OS Synthetic.

OS Chimeric.

XX Key Location/Qualifiers

FT misc_RNA 243..1021

FT /tag= a

FT /label= udp_gene

FT misc_RNA 1037..1766

FT /tag= b

FT /label= deoD_gene

FT misc_RNA 2229..3628

FT /tag= c

FT /label= pBR322_Tet_gene_sequence

XX WO200039307-A2.

XX 06-JUL-2000.

XX 23-DEC-1999; 99WO-EP010416.

XX 23-DEC-1998; 98IT-MI002792.

PR (NORP-) NORPHARMA SPA.

PA Bestetti G, Calli' S, Ghisotti D, Orsini G, Tonon G, Zuffi G;

XX WPI; 2000-452402/39.

XX Recombinant expression vectors used to transform cells for the production

PT of e.g. nucleosides encode uridine phosphorylase and/or purine nucleoside

PT phosphorylase.

XX Claim 14; Page 53-55; 72pp; English.

XX Host cells genetically modified to express uridine phosphorylase (Udp)

CC and purine nucleoside phosphorylase (PNP) or their corresponding crude or

CC purified extracts, either separately or in combination are used as

CC catalysts of transglycosylation reactions between a donor nucleoside and

CC an acceptor base, for preparing nucleoside analogues containing

CC heterocyclic systems with purine and/or pyrimidine bases substituted by

CC one or more nitrogen atoms, for preparing alpha-pentose-1-phosphate

CC sugars by phosphorolysis reactions and for producing nucleosides and

CC modified nucleoside analogues (all claimed). The modified or natural

CC nucleosides are used directly or as intermediates in the preparation of

CC drugs with anti-viral and anti-tumour activity and for preparing

CC oligonucleotides for therapeutic or diagnostic use. The recombinant host

CC cells are produced using recombinant plasmid expression vectors

CC comprising at least one gene sequence of a mesophilic bacterium coding

CC for a polypeptide with udp activity and/or enzyme PNP activity and at

CC least one gene sequence coding for antibiotic resistance. In particular,

CC the E. coli udp and deoD genes are used. Recombinant strains produced

CC using the vectors express polypeptides with enzyme Udp and PNP activity

CC in large amounts, e.g. 340-1040 times higher Udp activity and 120-200

CC times higher PNP activity than non-transformed corresponding wild type

CC strains

XX SQ Sequence 6301 BP; 1507 A; 1625 C; 1686 G; 1483 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	4,89e-138	Length:	6301
Score:	1222.00	Matches:	238
Percent Similarity:	100.00%	Conservative:	1
Best Local Similarity:	99.58%	Mismatches:	0
Query Match:	99.76%	Indels:	0

DB: 3 Gaps: 0

US-10-035-300A-2 (1-239) x AAAS1635 (1-6301)

QY 1 MetAlaThrProHisIleAsnAlaGluMetGlyAspPheAlaAspValValLeuMetPro 20

DB 1046 ATGGCTACCCACACATTAATGCAGAAATGGCGATTTCGCTGACGTAGTTTGTATGCCA 1105

QY 21 GlyAspProLeuArgAlaLysIleAlaGluThrPheLeuGluAspAlaArgGluVal 40

DB 1106 GCGACCCCGCTCGGTGCGAAGTATATCTGTAACCTTCTTGAAGATGCCGCTGAAGTG 1165

QY 41 AsnAsnValArgGlyMetLeuGlyPheThrGlyThrLysGlyArgLysIleSerVal 60

DB 1166 AACAACTTCGCGGTATGCTGGGCTTCACCGGTACTTACAAAGGCCGCAAAATTTCCGTA 1225

QY 61 MetGlyHisGlyValGlyIleProSerCysSerIleTyrThrLysGluLeuIleThrAsp 80

DB 1226 ATGGGTACCGGTATGGGTATCCCGTCTCTGCTCATCAACCAAGAACTGATACCCGAT 1285

QY 81 PheGlyValLysLysIleIleArgValGlySerCysGlyAlaValLeuProHisValLys 100

DB 1286 TTGCGGTGAAGAAATTAATCCGCGTGGGTTCTGTGGCGAGTTCTCGCGCACGTAAA 1345

QY 101 LeuArgAspValLysIleGlyMetGlyAlaCysThrAspSerLysValAsnArgIleArg 120

DB 1346 CTGCGCGACGTCGTTATCGGTATGGGTGCTGACCGATTCCAAAGTTAAACCGCATCGGT 1405

QY 121 PheLysAspHisAspPheAlaIleAlaAspPheAspMetValArgAsnAlaValAsp 140

DB 1406 TTTAAAGACCATGACTTTGCGCGCTATCGCTGACTTTCGACATGGTGGCGTAACCGCATGAT 1465

QY 141 AlaAlaLysAlaLeuGlyIleAspAlaArgValGlyAsnLeuPheSerAlaAspLeuPhe 160

DB 1466 GCAGCTAAAGCACTGGGTATTGATGCTGCGTGGGTAACTGTTCTCGCTGACCTGTC 1525

QY 161 TyrSerProAspGlyGluMetPheAspValMetGluLysTyrGlyIleLeuGlyValGlu 180

DB 1526 TACTCTCGGACGGCGAAATGTTGACGCTGATGGAATAATACGGCATTTCTCGGCGTGGAA 1585

QY 181 MetGluAlaAlaGlyIleTyrGlyValAlaAlaGluPheGlyAlaLysAlaLeuThrIle 200

DB 1586 ATGGAAGCGGTGGTATCTACGCGCTGCTGAGAAATTTGGCGGAAGCCCTGACCATC 1645

QY 201 CysThrValSerAspHisIleArgThrHisGluGlnThrThrAlaAlaGluArgGlnThr 220

DB 1646 TGCACCGTATTGACCATTCGACATCCGACATCAGCAGCAGACCATCGCGCTGAGCGTCAGACT 1705

QY 221 ThrPheAsnAspMetIleLysIleAlaLeuGluSerValLeuLeuGlyAspLysGlu 239

DB 1706 ACCTTCAACGACATGATCAAAATCGCACTGGAATCCGTTCTGCTGGCGGATAAAGAG 1762

RESULT 13

AAD56238

ID AAD56238 standard; DNA; 720 BP.

XX AAD56238;

AC AAD56238;

XX 07-AUG-2003 (first entry)

XX Escherichia coli mutant PNP enzyme (A157V) encoding DNA #SEQ ID 3.

DE Purine nucleoside phosphorylase; mutant; tumour; PNP; enzyme; cancer;

XX leukaemia; glioblastoma; lymphoma; melanoma; carcinoma; gene therapy;

KW gene; ds.

KW Escherichia coli.

OS Synthetic.

XX Key Location/Qualifiers

FT CDS 1..720

FT /*tag= a

FT /product= "Mutant PNP enzyme"

CC The present sequence encodes a purine nucleoside phosphorylase enzyme.
CC This enzyme is involved in the biosynthesis of deoxyribonucleosides, and
CC is used in the invention. The specification describes a
CC method for the in vitro enzymatic synthesis of deoxyribonucleosides. The
CC method comprises reacting deoxyribose 1-phosphate and a nucleobase to
CC form a deoxyribonucleoside and an inorganic phosphate. Enzymes which may
CC be used in the method of the invention include thymidine phosphorylase,
CC purine nucleoside phosphorylase, phosphopentose mutase, phosphopentose
CC aldolase, fructose 1,6-diphosphate aldolase, deoxyribokinase, and
CC nucleoside 2-deoxyribosyltransferase
XX

SQ Sequence 720 BP; 174 A; 190 C; 185 G; 171 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 3.08e-138 Length: 720
Score: 1213.00 Matches: 236
Percent Similarity: 99.58% Conservativity: 2
Best Local Similarity: 98.74% Mismatches: 1
Query Match: 99.02% Indels: 0
DB: 5 Gaps: 0

US-10-035-300A-2 (1-239) x AAF55446 (1-720)

Qy 1 MetAlaThrProHisIleAsnAlaGluMetGlyAspPheAlaAspValValLeuMetPro 20
Db 1 ATGGCTACCCACACATTAAATGCGAATGGCGATTTCGCTGACGTAGTTTATGCGCA 60
Qy 21 GlyAspProLeuArgAlaLysTyrIleAlaGluThrPheLeuGluAspAlaArgGluVal 40
Db 61 GCGACCGCGCTGCGCGAAGTATATCTGCTGAACTTCTTGAAGATGCCCGTGAAGTG 120
Qy 41 AsnAsnValArgGlyMetLeuGlyPheThrGlyThrTyrLysGlyArgLysIleSerVal 60
Db 121 AACACGTTCCGCGTATCTCGGCTTACCGGTACTTACAAAGCGCGCAAAATTTCCGTA 180
Qy 61 MetGlyHisGlyValGlyIleProSerCysSerIleTyrThrLysGluLeuThrAsp 80
Db 181 ATGGGTACGGTATGGGTATCCCGTCTGCTCCATCTACCAAGAACTGATCACCAGT 240
Qy 81 PheGlyValLysLysIleIleArgValGlySerCysGlyAlaValLeuProHisValLys 100
Db 241 TTCGCGTGAAGAAATATATCGCGTGGTTCCTGTGGCGAGTTCTCGCGCAGTAA 300
Qy 101 LeuArgAspValValIleGlyMetGlyAlaCysThrAspSerLysValAsnArgIleArg 120
Db 301 CTGCGCGCAGTCGTATCGGTATGGGTACCTGCACCGATTCCAAAGTTTAAACCGCATCGT 360
Qy 121 PheLysAspHisAspPheAlaIleAlaAspPheAspMetValArgAsnAlaValAsp 140
Db 361 TTTAAGACCATGACTTTGCGGCTATCGCTGACTTCGACATGGTGGCGTAACGCGTAGAT 420
Qy 141 AlaAlaLysAlaLeuGlyIleAspAlaArgValGlyAsnLeuPheSerAlaAspLeuPhe 160
Db 421 GCAGCTAAAGCACTGGGTATGTATGCTGCGTGGGTAACTGTCTCGCTGACCTGTC 480
Qy 161 TyrSerProAspGlyGluMetPheAspValMetGluLysTyrGlyIleLeuGlyValGlu 180
Db 481 TACTCTCCGACGGCGAAATGTTGACGTGATGGAATAACGGCATTTCTCGCGGTGAA 540
Qy 181 MetGluAlaAlaGlyIleTyrGlyValAlaAlaGluPheGlyAlaLysAlaLeuThrIle 200
Db 541 ATGAAGCGGCTGTGATCTACGGGCTCGCTCAGAAATTTGCGCGAAGCCCTGACCATC 600
Qy 201 CysThrValSerAspHisIleArgThrHisGluGlnThrThrAlaAlaGluArgGlnThr 220
Db 601 TGCACCGTATCTGACCACATCCGACTCAGCAGGACGACCATCTCGCTGACGCTGACACT 660
Qy 221 ThrPheAsnAspMetIleLysIleAlaLeuGluSerValLeuLeuGlyAspLysGlu 239
Db 661 ACCTTCAACACATGATCAAAATCGCACTGGAATCCGTTCTGCTGGGCGATAAAGAG 717

RESULT 15
ACH95836

ID ACH95836 standard; DNA; 732 BP.

XX ACH95836;

XX 29-JUL-2004 (first entry)

XX Klebsiella pneumoniae polynucleotide seqid 1631.

XX Recombinant expression vector; transcription regulatory element;
KW Klebsiella pneumoniae protein; antibacterial; vaccine; gene; ds.

XX Klebsiella pneumoniae.

XX US6610836-B1.

XX 26-AUG-2003.

XX 27-JAN-2000; 2000US-00489039.

XX 29-JAN-1999; 99US-0117747P.

XX (GENO-) GENOME THERAPEUTICS CORP.

XX Breton GL, Osborne M;

XX WPI; 2003-895346/82.

XX P-ESDB; ABO62285.

XX New nucleic acid encoding a Klebsiella pneumoniae polypeptide, useful for
preparing a vaccine composition against Klebsiella pneumoniae.

XX Disclosure; SEQ ID NO 1631; 932pp; English.

XX The invention describes a new isolated nucleic acid encoding a Klebsiella
pneumoniae polypeptide. Also described are: a recombinant expression
vector comprising the nucleic acid, operably linked to a transcription
regulatory element; and a cell comprising the recombinant expression
vector. The nucleic acid is useful for preparing a vaccine composition
against Klebsiella pneumoniae. This sequence encodes a Klebsiella
pneumoniae polypeptide of the invention

XX Sequence 732 BP; 167 A; 200 C; 221 G; 144 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 5.33e-132 Length: 732
Score: 1162.00 Matches: 226
Percent Similarity: 97.07% Conservativity: 6
Best Local Similarity: 94.56% Mismatches: 7
Query Match: 94.86% Indels: 0
DB: 11 Gaps: 0

US-10-035-300A-2 (1-239) x ACH95836 (1-732)

Qy 1 MetAlaThrProHisIleAsnAlaGluMetGlyAspPheAlaAspValValLeuMetPro 20
Db 13 ATGCACTCTCTACATTAACGCGAAGTTCGCTGACGTGCTTTTATGCGG 72
Qy 21 GlyAspProLeuArgAlaLysTyrIleAlaGluThrPheLeuGluAspAlaArgGluVal 40
Db 73 GCGACCGCTGCGCGGAGACATTTGCAGAACTTCTCGAAGACGTCGCTGAAGTG 132
Qy 41 AsnAsnValArgGlyMetLeuGlyPheThrGlyThrTyrLysGlyArgLysIleSerVal 60
Db 133 AACACGTCGCGCATGCTGGGCTTCCCGGTACCTATTAAGCCGCAAAATCTCCG 192
Qy 61 MetGlyHisGlyValGlyIleProSerCysSerIleTyrThrLysGluLeuThrAsp 80
Db 193 ATGGGCGACGGATGGGGATCCCGTCTGCTATCTACACCAAGAGCTGATCACC 252
Qy 81 PheGlyValLysLysIleIleArgValGlySerCysGlyAlaValLeuProHisValLys 100
Db 253 TTCGCGTGAAGAAATATCTCCGCTGGGCTCTTCGCGCGGCTGCGTGAAGATGTC 312

```
Qy 101 LeuArgAspValValIleGlyMetGlyAlaCysThrAspSerLysValAsnArgIleArg 120
Db |||||
Db 313 CTGGCTGACGTGGTGAATCGGCATGGGGCGGTGCACCGACTCTAAAGTGAACCGTCTGCGT 372
|||
Qy 121 PheLysAspHisAspPheAlaIleAlaAspPheAspMetValArgAsnAlaValAsp 140
Db |||||
Db 373 TTCAAAGACCATGATTTTGGCGCATTCGCCGATTTCCGCATGGTGGCTAATGGCGTAGAC 432
|||
Qy 141 AlaAlaLysAlaLeuGlyIleAspAlaArgValGlyAsnLeuPheSerAlaAspLeuPhe 160
Db |||||
Db 433 GCGGCGAAAGCGCTGGGCGTGTAGCGCGCGGTGCGCAACATCTTCTCCGAGACCTGTTC 492
|||
Qy 161 TyrSerProAspGlyGluMetPheAspValMetGluLysTyrGlyIleLeuGlyValGlu 180
Db |||||
Db 493 TATAGCGCGGACCGCTCATGTTTCGACGTGATGGAAAAATACGGCATTCGGGCGTGGAA 552
|||
Qy 181 MetGluAlaAlaGlyIleTyrGlyValAlaAlaGluPheGlyAlaLysAlaLeuThrIle 200
Db |||||
Db 553 ATGGAAGCGCGGTATCTACGGCGTGGCGGAGTTCGGCGCGAAGCGCTGACCATC 612
|||
Qy 201 CysThrValSerAspHisIleArgThrHisGluGlnThrThrAlaAlaGluArgGlnThr 220
Db |||||
Db 613 TGCACCGTCTCCGACCATCCGTACCCAGCAGCAGACCACTGCCGCTGAACGTCAGACG 672
|||
Qy 221 ThrPheAsnAspMetIleLysIleAlaLeuGluSerValLeuLeuGlyAspLysGlu 239
Db |||||
Db 673 ACCTTCACACGACATGATCAAAATCGCGCTGGATCCGTGCTGCTGGGCGATRAAGAG 729
|||
```

Search completed: July 23, 2005, 01:51:40
Job time : 688 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: July 23, 2005, 01:24:48 ; Search time 220 Seconds
(without alignments)
1777.592 Million cell updates/sec

Title: US-10-035-300A-2

Perfect score: 1225

Sequence: 1 MATPHINAEMGDFADVLLMP.....TTFNDMIKIALESVLLGDKE 239

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delep 6.0 , Delext 7.0

Searched: 1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO.spool_p/US10035300/runat_18072005_165351_28002/app.query.fasta_1.391
-DB=Issued_Patents_NA -QPMT=fastap -SUFFIX=rni -MINMATCH=0.1 -LOOPCL=0
-LOOPTXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blomsum62 -TRANS=human40.cdi
-LIST=45 -DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pct -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10035300 @CIGN 1 1 105 @runat_18072005_165351_28002 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEOQUERY -NEG_SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAEXT=0.5 -FGAPOP=6
-FGAEXT=7 -YGAPOP=10 -YGAEXT=0.5 -DELOP=6 -DELEXT=7

Database :

- 1: /cgn2_6/ptodata/1/ina/5A.COMB.seq.*
- 2: /cgn2_6/ptodata/1/ina/5B.COMB.seq.*
- 3: /cgn2_6/ptodata/1/ina/6A.COMB.seq.*
- 4: /cgn2_6/ptodata/1/ina/6B.COMB.seq.*
- 5: /cgn2_6/ptodata/1/ina/PCTUS.COMB.seq.*
- 6: /cgn2_6/ptodata/1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1222	99.8	5013	US-09-183-188B-5	Sequence 5, Appli
2	1162	94.9	732	US-09-489-039A-1631	Sequence 1631, Ap
3	1057	86.3	795	US-09-543-681A-2371	Sequence 2371, Ap
C 4	1013	82.7	1830121	US-09-557-884-1	Sequence 1, Appli
C 5	1013	82.7	1830121	US-09-643-990A-1	Sequence 1, Appli
C 6	778	63.5	640681	US-09-790-988-1	Sequence 1, Appli
7	728	59.4	744	US-09-902-540-3568	Sequence 3568, Ap
C 8	728	59.4	19019	US-09-902-540-1171	Sequence 1171, Ap
9	607	49.6	735	US-09-107-532A-1185	Sequence 1185, Ap
10	591	48.2	714	US-09-134-000C-2608	Sequence 2608, Ap
11	576.5	47.1	711	US-09-583-110-37	Sequence 37, Appl
12	576.5	47.1	714	US-09-107-433-1967	Sequence 1967, Ap

13	575.5	47.0	13440	3	US-08-961-527-128	Sequence 128, App
14	572	46.7	732	3	US-09-134-001C-1402	Sequence 1402, Ap
15	569	46.4	2962	4	US-09-710-279-3511	Sequence 3511, Ap
16	569	46.4	5301	4	US-08-956-171B-443	Sequence 443, App
17	569	46.4	5301	4	US-08-781-986A-443	Sequence 443, App
18	557.5	45.5	2006	4	US-08-956-171B-378	Sequence 378, App
19	557.5	45.5	2006	4	US-08-781-986A-378	Sequence 378, App
C 20	538.5	44.0	5739	4	US-09-634-238-1	Sequence 1, Appli
21	466	38.0	580073	4	US-08-545-528D-1	Sequence 1, Appli
22	300	24.5	1124	4	US-08-809-254A-3	Sequence 3, Appli
23	245.5	20.0	897	4	US-09-489-039A-860	Sequence 860, App
24	243.5	19.9	22108	3	US-09-053-197A-3	Sequence 3, Appli
25	243.5	19.9	22108	3	US-09-085-761A-3	Sequence 3, Appli
26	231	18.9	1830121	4	US-09-557-884-1	Sequence 1, Appli
27	231	18.9	1830121	4	US-09-643-990A-1	Sequence 1, Appli
28	227.5	18.6	798	4	US-09-543-681A-1537	Sequence 1537, Ap
29	167	13.6	3072	3	US-09-221-017B-337	Sequence 337, App
30	139	11.3	711	4	US-09-107-532A-3162	Sequence 3162, Ap
31	136	11.1	717	4	US-09-489-039A-5257	Sequence 5257, Ap
32	131	10.7	729	4	US-09-543-681A-1894	Sequence 1894, Ap
33	130	10.6	699	3	US-09-347-878-17	Sequence 17, Appl
34	130	10.6	699	4	US-09-711-164-226	Sequence 226, App
35	122.5	10.0	711	4	US-09-134-000C-1699	Sequence 1699, Ap
36	116.5	9.5	927	4	US-09-328-352-534	Sequence 534, App
37	115	9.4	687	4	US-09-710-279-2935	Sequence 2935, Ap
38	115	9.4	687	4	US-09-710-279-3195	Sequence 3195, Ap
39	115	9.4	702	3	US-09-134-001C-806	Sequence 806, App
C 40	115	9.4	3781	4	US-09-710-279-3993	Sequence 3993, Ap
41	115	9.4	3805	4	US-09-710-279-3687	Sequence 3687, Ap
42	110	9.0	4104	3	US-08-961-527-169	Sequence 169, App
43	108	8.8	693	4	US-09-583-110-2551	Sequence 2551, Ap
44	108	8.8	720	4	US-09-107-433-826	Sequence 826, App
45	106	8.7	765	4	US-09-583-110-2358	Sequence 2358, Ap

ALIGNMENTS

RESULT 1

US-09-183-188B-5
; Sequence 5, Application US/09183188B
; Patent No. 6491905
; GENERAL INFORMATION:
; APPLICANT: Sorscher, Eric J.
; APPLICANT: Parker, William B.
; APPLICANT: Gadi, Vijaykrishna K.
; TITLE OF INVENTION: RECOMBINANT E. COLI FOR DELIVERY OF PNP TO TUMOR CELLS
; FILE REFERENCE: UAB-12405/22
; CURRENT APPLICATION NUMBER: US/09/183,188B
; PRIOR FILING DATE: 1998-10-30
; PRIOR FILING DATE: 1996-08-23
; PRIOR FILING DATE: 1993-09-14
; PRIOR FILING DATE: 1997-10-31
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5
; LENGTH: 5013
; TYPE: DNA
; ORGANISM: Escherichia coli
US-09-183-188B-5
Alignment Scores:
Pred. No.: 5.28e-154
Score: 1222.00
Percent Similarity: 100.00%
Best Local Similarity: 99.58%
Query Match: 99.76%
DB: 4
Length: 5013
Matches: 238
Conservative: 1
Mismatches: 0
Indels: 0
Gaps: 0
US-10-035-300A-2 (1-239) x US-09-183-188B-5 (1-5013)

```
Qy 1 MetAlaThrProHisIleAsnAlaGluMetGlyAspPheAlaAspValValLeuMetPro 20
Db 413 ATGGCTACCCACACATTAATGCAAGAAATGGCGATTTTCGTGACGAGTTTGTATGCCA 472

Qy 21 GlyAspProLeuArgAlaIleAsnAlaGluThrPheLeuGluAspAlaArgGluVal 40
Db 473 GCGACCCGCTGCGTGGAGTATATGCTGAAACTTTCTTGAAGATGCCCGTGAAGTG 532

Qy 41 AsnAsnValArgGlyMetLeuGlyPheThrGlyThrTyrlsGlyArgIleSerVal 60
Db 533 AACAACTGTCGCGTATGCTGGCTTCCACCGTACTTACAAAGCCGCCAAAATTTCCGTA 592

Qy 61 MetGlyHisGlyValGlyIleProSerCysSerIleTyrlsGlyLeuLeuIleThrAsp 80
Db 593 ATGGGTCAACGGTATGGGTATCCCGTCTCTGCTCATCTACACCAAGAACTGATCACCGAT 652

Qy 81 PheGlyValIleValIleIleArgValGlySerCysGlyAlaValLeuProHisValIys 100
Db 653 TTCGGCGTGAAGAAATATATCCGGTGGTTCCTGTGGCGAGTTCTGCGCACGTAAAA 712

Qy 101 LeuArgAspValValIleGlyMetGlyAlaCysThrAspSerIysValAsnArgIleArg 120
Db 713 CTGCGCACGTCGTTATCGGTATGGGTGCTGACCGATTTCCAAAGTTAAACCGCATCCGT 772

Qy 121 PheIleAspHisAspPheAlaAlaIleAlaAspPheAspMetValArgAsnAlaValAsp 140
Db 773 TTTAAAGACCATGACTTTGGCGCTATCGCTGATCTGACATGTTGCGTAACGCGATAGAT 832

Qy 141 AlaAlaIysAlaLeuGlyIleAspAlaArgValGlyAsnLeuPheSerAlaAspLeuPhe 160
Db 833 GCAGCTAAAGCAGCTGGTATTGATGCTCGCGTGGGTAACCTGTTCTCCGCTGACCTGTTT 892

Qy 161 TyrSerProAspGlyGluMetPheAspValMetGlyIleValGlyIleLeuGlyValGlu 180
Db 893 TACTCTCCGACGCGGAAATGTTTCGACGTGATGGAAATAACGCGATTTCTCGCGGTGGAA 952

Qy 181 MetGluAlaAlaGlyIleTyrlsGlyValAlaAlaGluPheGlyAlaIysAlaLeuThrIle 200
Db 953 ATGGAAACGCGTGTATCTACGCGCTCGCTGCAGAAATTTGGCGGAAGCCCTGCACATC 1012

Qy 201 CysThrValSerAspHisIleArgThrHisGluGlnThrThrAlaAlaGluArgGlnThr 220
Db 1013 TGCACCGTATCTGACACATCCGCACTACGAGCAGACCACTGCGCGTGAAGCTCAGACT 1072

Qy 221 ThrPheAsnAspMetIleIysIleAlaLeuGluSerValLeuLeuGlyAspIysGlu 239
Db 1073 ACCTTCAACGACATGATCAAAATCGCACTGGAATCCGTTCTGCTGGGCGATAAAGAG 1129

RESULT 2
US-09-489-039A-1631
; Sequence 1631, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489.039A
; PRIOR FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 1631
; LENGTH: 732
; TYPE: DNA
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-1631

Alignment Scores:
Pred. No.: 3.48e-147 Length: 732
Score: 1162.00 Matches: 226
Percent Similarity: 97.07% Conservative: 6
```

```
Best Local Similarity: 94.56% Mismatches: 7
Query Match: 94.86% Indels: 0
DB: 4 Gaps: 0

US-10-035-300A-2 (1-239) x US-09-489-039A-1631 (1-732)

Qy 1 MetAlaThrProHisIleAsnAlaGluMetGlyAspPheAlaAspValValLeuMetPro 20
Db 13 ATGGCAACTCTCACATTAACGCAAGAAATGGCGATTTTCGTGACGTCGTTTGTATGCCG 72

Qy 21 GlyAspProLeuArgAlaIleAsnAlaGluThrPheLeuGluAspAlaArgGluVal 40
Db 73 GCGACCCGCTGCGCGCAAGCAATTCGAGAAACCTTCTCGAAGACGTCGTCGAAGTG 132

Qy 41 AsnAsnValArgGlyMetLeuGlyPheThrGlyThrTyrlsGlyArgIleSerVal 60
Db 133 AACAACTGCGCGCATGCTGGCGTTTCACCGGTACCTATAAAGCGCGCAAAAATCTCCGTG 192

Qy 61 MetGlyHisGlyValGlyIleProSerCysSerIleTyrlsGlyLeuLeuIleThrAsp 80
Db 193 ATGGCGCACGGGATGGGATCCGCTCTGCTCTATCTACACCAAGAGCTGATCACCGAT 252

Qy 81 PheGlyValIysIysIleIleArgValGlySerCysGlyAlaValLeuProHisValIys 100
Db 253 TTCGGCGTGAAGAAATCATCCGCTGGGCTCTCGCGCGCGTGGCGTGAAGATGTCAAA 312

Qy 101 LeuArgAspValValIleGlyMetGlyAlaCysThrAspSerIysValAsnArgIleArg 120
Db 313 CTGCGTGACGTGTGATCGGCATGGCGCGGTGACCGACTCTTAAAGTGAACCGCTGCGGT 372

Qy 121 PheIleAspHisAspPheAlaAlaIleAlaAspPheAspMetValArgAsnAlaValAsp 140
Db 373 TTCAAAGACCATGATTTTGGCGCATTTGCCGATTTGCCGATGTTGCGTAAATGCGTAGAC 432

Qy 141 AlaAlaIysAlaLeuGlyIleAspAlaArgValGlyAsnLeuPheSerAlaAspLeuPhe 160
Db 433 GCGCGAAAACGCTGGCGGTGACGCGCGCTGCGCAACATCTTCTCCGCGACACCTGTTC 492

Qy 161 TyrSerProAspGlyGluMetPheAspValMetGlyIleTyrlsGlyIleLeuGlyValGlu 180
Db 493 TATACCGCGACCCGCTCCATGTTTCGACGTGATGGAAATAATACCGCATTTGCGCGTGGAA 552

Qy 181 MetGluAlaAlaGlyIleTyrlsGlyValAlaAlaGluPheGlyAlaIysAlaLeuThrIle 200
Db 553 ATGGAAGCGCGGTATCTACGCGCTGGCGCGAGTTCGCGCGCAAGCGCTGACCATC 612

Qy 201 CysThrValSerAspHisIleArgThrHisGluGlnThrThrAlaAlaGluArgGlnThr 220
Db 613 TGCACCGTCTCCGACCATCATCCGTATCCACGACGACCACTGCGCGTGAACGTCAGACG 672

Qy 221 ThrPheAsnAspMetIleIysIleAlaLeuGluSerValLeuLeuGlyAspIysGlu 239
Db 673 ACCTTCAACGACATGATCAAAATCGCGTGGAAATCCGCTGCTGCGCGGATAAAGAG 729

RESULT 3
US-09-543-681A-2371
; Sequence 2371, Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION:
; APPLICANT: GARY BRETON
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.1002-001
; CURRENT APPLICATION NUMBER: US/09/543.681A
; CURRENT FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: US 60/128,706
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 2371
; LENGTH: 795
; TYPE: DNA
; ORGANISM: Proteus mirabilis
US-09-543-681A-2371
```

```
Alignment Scores:
Pred. No.: 6.07e-133 Length: 795
Score: 1057.00 Matches: 205
Percent Similarity: 93.62% Conservative: 15
Best Local Similarity: 87.23% Mismatches: 15
Query Match: 86.29% Indels: 0
DB: 4 Gaps: 0

US-10-035-300A-2 (1-239) x US-09-543-681A-2371 (1-795)

Qy 1 MetAlaThrProHisIleAsnAlaGluMetGlyAspPheAlaAspValValLeuMetPro 20
Db 79 ATGGCTACCCCTCATATTAAACGCGAAGATGGCGATTTTGTGATGTCGTTTAAATGCGG 138
Qy 21 GlyAspProLeuArgAlaLysIleAlaGluThrPheLeuGluAspAlaArgGluVal 40
Db 139 GCGACCGCGCTCGTGTCTAATAATCAATGCTGAAACTTTTACAGAGCTACGTCAGTA 198
Qy 41 AsnAsnValArgGlyMetLeuGlyPheThrGlyThrIleValArgLysIleSerVal 60
Db 199 AATAATGTTCTGGCATGTTAGGTTTACTGCTACGTATAAAGCGGTAAATTTCTGTA 258
Qy 61 MetGlyHisGlyValGlyIleProSerCysSerIleThrIleValLeuThrAsp 80
Db 259 ATGGGCCACCGTATGGGTATTTCTTCTGCTCAATTTATGCTAAGAGCTAATTACTGAT 318
Qy 81 PheGlyValLysLysIleAlaArgValGlySerCysGlyAlaValLeuProHisValLys 100
Db 319 TTGGTGTAAGCAAGTATTTCCGCTGGGTTCATGTTGGTGGGTATTTACCAAGATGTTGNA 378
Qy 101 LeuArgAspValValIleGlyMetGlyAlaCysThrAspSerLysValAsnArgIleArg 120
Db 379 TTACGTGATGTTGTGATCGGTATGGGACCATGTACAGATTTCTAAAGTTTAAACGGCTTACG 438
Qy 121 PheLysAspHisAspPheAlaAlaIleAlaAspPheAspMetValArgAsnAlaValAsp 140
Db 439 TTTAAAGACCAAGATTTTTCGGCGCTATTGCTGATTTTGAATTTAGTACAAATGCTGTATCT 498
Qy 141 AlaAlaLysAlaLeuGlyIleAspAlaArgValGlyAsnLeuPheSerAlaAspLeuPhe 160
Db 499 GCTCGGAAAGCAAGATATTAAAGTTCGTGGGTATATATTTCTCAGCGGATCTTTTC 558
Qy 161 TyrSerProAspGlyGluMetPheAspValMetGluLysIleValLeuGlyValGlu 180
Db 559 TACTCGCTGATCCAGAAATGTTGATGTGATGCGAAATACCGGCATCTTAGGTGTTGAA 618
Qy 181 MetGluAlaAlaGlyIleThrGlyValAlaAlaGluPheGlyAlaLysAlaLeuThrIle 200
Db 619 ATGGAAGCAGCAGGATTTATGTTGATGCTGCGGAATATGTCACGTCGCCCTGACAAAT 678
Qy 201 CysThrValSerAspHisIleArgThrHisGluGlnThrThrAlaAlaArgGlnThr 220
Db 679 TGTACTGTTTCTGATCATATTAAGGAGGTACAAACACCTCAGAGAGCGCCAAACT 738
Qy 221 ThrPheAsnAspMetIleLysIleAlaLeuGluSerValLeuLeu 235
Db 739 ACCTTTAATGAATGATTGAATTCATAGCAATCTGTTCTGTTA 783

RESULT 4
US-09-557-884-1/c
; Sequence 1, Application US/09557884
; Patent No. 6506581
; GENERAL INFORMATION:
; APPLICANT: Fleischmann et al.
; TITLE OF INVENTION: The Nucleotide sequence of
; the Haemophilus influenzae Rd Genome, Fragments
; Thereof, and Uses Thereof
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
```

Qy	163	ProAspGlyGluMetPheAspValMetGlnLysTyrGlyIleLeuGlyValGluMetGlu	182
Db	540396	CCCCATGTGGAAATGTTCCATGTATGTGAAAAATACCGCATTTTAGGTGTGGAATGGAA	540337
Qy	183	AlaalaGlyIleTyrGlyValAlaalaGluPheGlyAlaIleAlaLeuThrIleCysThr	202
Db	540336	GCTCCTGGAATTTATGGTGTGGCTGCAGAAATATGGTGCAAAAGCAATTAAACGATTTTGTACC	540277
Qy	203	ValSerAspHisIleArgThrHisGluGlnThrThrAlaAlaGlyArgGlnThrThrPhe	222
Db	540276	GTGTCTGACCATATTCGTACTACTACGAACAAACCAGCGGAGAACGTCATTAACATTC	540217
Qy	223	AsnAspMetIleLysIleAlaLeuGlnSerValLeuLeuGlyAsp	237
Db	540216	AATCATATGATTGAAATTCGCTAGATTCAGTATTAAATGGTGAT	540172

DB: 4 Gaps: 0

US-10-035-300A-2 (1-239) x US-09-643-990A-1 (1-1830121)

Qy 3 ThrProHisIleAsnAlaGluMetGlyAspPheAlaAspValValLeuMetProGlyAsp 22
 Db 540876 ACTCCACATATTAAACGGCCTGAAGGTGCATTTCTGCTGATGTGTTTAATCCCTGGCCGAT 540817

Qy 23 ProLeuArgAlaIlystYrIleAlaGluThrPheLeuGluAspAlaAaGluValAsnAsn 42
 Db 540816 CCATCTCGTCGCAAAATATATTGCAGAACTTTTTCACAGATGTGTGTGAAGTAAACGAAT 540757

Qy 43 ValArgGlyMetLeuGlyPheThrGlyThrTyrlsGlyArgIlyslleSerValMetGly 62
 Db 540756 GTTCGTAATATGCTTGGTTTTACTGGAACTTATAAAGCTGTAAATTTCTATTATGGGG 540697

Qy 63 HisGlyValGlyIleProSerCysSerIleTyThrIysGluLeuIleThrAspPheGly 82
 Db 540696 CACGGTATGGGATTCATCTTCTATTACCGCAAGAATAAATCAGTAAATATGGC 540633

Qy 83 ValIlyslsIleIleAaGValGlySerCysGlyAlaValLeuProHisValIysLeuArg 102
 Db 540636 GTGAAAAAATATCCGTGTAGGTTCTGTGGCACTGTCGTATGGACGAGTGAAGTGGCC 540577

Qy 103 AspValValIleGlyMetGlyAlaCysThrAspSerIysValAsnAaGilleArgPheIys 122
 Db 540576 GATGTGATTATCGGTTTAGGTGCATGACTATCAAAAGTAAACCGTATTTCGTTTCAAA 540517

Qy 123 AspHisAspPheAlaAlaIleAlaAspPheAspMetValArgAsnAlaValAspAlaAla 142
 Db 540516 GATAACAGACTTTCAGCTATTTGCTGACTTTTGATATGGCACAAAGCCGCTGTTCAAGCTGCA 540457

Qy 143 LysAlaLeuGlyIleAspAlaAaGValGlyAsnLeuPheSerAlaAspLeuPheTySer 162
 Db 540456 AAAGCAAAAGTAAAGTCGTCGTGTAGGTAATTTATCTCGCGGATTTATTTCTATACA 540397

Qy 163 ProAspGlyGluMetPheAspValMetGluIysTyrlleLeuGlyValGluMetGlu 182
 Db 540396 CCCGATGTGAAATGTTTCGATGTAATGGAAAAATACGGCATTTTAGGTGCGAAATGGAA 540337

Qy 183 AlaAlaGlyIleTyrlleGlyValAlaAlaGluPheGlyAlaIysAlaLeuThrIleCysThr 202
 Db 540336 GCTGCTGGAATTTATGCTGTGGCTGCGAGATATGTCGCAAGCATTTAAACGATTTGTACC 540277

Qy 203 ValSerAspHisIleAaGThrHisGluGlnThrThrAlaAlaGluAaGlnThrThrPhe 222
 Db 540276 GTGTCTGACCATATTCTGATCTCAGAAACAAACACGCGCAGNAGAACGTCATTAACATTC 540217

Qy 223 AsnAspMetIleIysIleAlaLeuGluSerValLeuLeuGlyAsp 237
 Db 540216 AATGATATGATTGAAATTCGTTAGATTAGATTAAATTCGTGAT 540172

RESULT 6

US-09-790-988-1/c

; Sequence 1, Application US/09790988

; Patent No. 6632935

; GENERAL INFORMATION:

; APPLICANT: SHIGENOBU, SHUJI

; APPLICANT: WATANABE, HIDEMI

; APPLICANT: HATTORI, MASAHIRA

; APPLICANT: SAKAKI, YOSHIYUKI

; TITLE OF INVENTION: GENOME DNA OF BACTERIAL SYMBIONT OF APHIDS

; FILE REFERENCE: 081356/0159

; CURRENT APPLICATION NUMBER: US/09/790,988

; CURRENT FILING DATE: 2001-02-23

; PRIOR APPLICATION NUMBER: JP2000-107160

; PRIOR FILING DATE: 2000-04-07

; NUMBER OF SEQ ID NOS: 7

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 1

; LENGTH: 640681

; TYPE: DNA

; ORGANISM: Buchnera sp.

US-09-790-988-1

Alignment Scores:

Pred. No.:	9.1e-91	Length:	640681
Score:	778.00	Matches:	144
Percent Similarity:	79.91%	Conservative:	43
Best Local Similarity:	61.54%	Mismatches:	47
Query Match:	63.51%	Indels:	0
DB:	4	Gaps:	0

US-10-035-300A-2 (1-239) x US-09-790-988-1 (1-640681)

Qy 1 MetAlaThrProHisIleAenAlaGluMetGlyAsePheAlaAaspValValLeuMetPro 20

Db 575671 ATGCTACTCCACATATTAATAGTAAAAAGATGATTTTTCAGATATATTTTAAATGCG 575612

Qy 21 GlyAseProLeuAargAlaIleAaGluThrPheLeuGluAaspAlaAargGluVal 40

Db 575611 GGAGCCAGCTTCGTGCAAAATATATGCTGAAAAATATTTTAAATATTTTTCAGTAA 575552

Qy 41 AsnAenValAargGlyMetLeuGlyPheThrGlyThrTyLysGlyAargLysIleSerVal 60

Db 575551 AATGATACCTCGTTAAATGTTAGCTTATACCGGATTTTATAAAAAATAGAAAGATTTCAATC 575492

Qy 61 MetGlyHisGlyValGlyIleProSerCysSerIleTyThrLysGluLeuIleThrAse 80

Db 575491 ATGAGTCATGATAGGAATACCATCAGCTTCTCTATACACAGAGATTAATATTGAA 575432

Qy 81 PheGlyValLysIleAaGluValGlySerCysGlyAlaValLeuProHisValLys 100

Db 575431 TTTAATGTAAAAAATATTTCTGTATAGGAACCTTGTGTGCTGATGATGATATAAG 575372

Qy 101 LeuAargAaspValValIleGlyMetGlyAlaCysThrAsePheSerLysValAasnAargIleAarg 120

Db 575371 TTACGTGATATAGTAAATCAGTATGGAGCTTCTACTGATTTCTAAAGTAAATAGATAAGA 575312

Qy 121 PheLysAseHisAaspPheAlaIleAaAaspPheAaspMetValAargAenAlaValAasp 140

Db 575311 TTTAATGATCATGACTTTGTGTCATATGACAGATTTTGTATGATGATTAATATATAGTCTCA 575252

Qy 141 AlaAalysAlaLeuGlyIleAaPheAaAargValGlyAasnLeuPheSerAlaAaspLeuPhe 160

Db 575251 ATTTCAAAAAATGAAAAATGCTCATTGGTAAATTTTTCACACAGATTTCTTTT 575192

Qy 161 TyrSerProAaspGlyGluMetPheAaspValMetGluLysTyLysGlyIleLeuGlyValGlu 180

Db 575191 TATAATGATGATAAAAAATGCTAAATATTTTAAAAAATATAATATATTATTCGAGTTGAC 575132

Qy 181 MetGluAlaAaGlyIleTyLysGlyAlaAaAaGluPheGlyAlaLysAlaLeuThrIle 200

Db 575131 ATGAGAGCTCGCGGAATATATGAGTGGCTTCTGAAATTTAAAAAGTTTCAACGATATATCGATA 575072

Qy 201 CysThrValSerAseHisIleAaGluThrHisGluGlnThrThrAlaAaAaGluAargGlnThr 220

Db 575071 TGTACAGTATCTCATATATTTACAAATAAAGATTTTCTTCATCAAAAGAAAGAGATCA 575012

Qy 221 ThrPheAasnAaspMetIleLysIleAlaLeuGluSerValLeu 234

Db 575011 AGTTTAAATGATGATGAGTACAGCTTGGATCTGTTTAA 574970

RESULT 7

US-09-902-540-3568

Sequence 3568, Application US/09902540

Patent No. 6833447

GENERAL INFORMATION:

APPLICANT: Goldman, Barry S.

APPLICANT: Hinkle, Gregory J.

APPLICANT: Slater, Steven C.

APPLICANT: Wiegand, Roger C.

TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof

FILE REFERENCE: 38-10(15849)B

CURRENT APPLICATION NUMBER: US/09/902,540

CURRENT FILING DATE: 2001-07-10


```
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wisegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15849)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; PRIORITY FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/217,883
; PRIOR FILING DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 1171
; LENGTH: 19019
; TYPE: DNA
; ORGANISM: Myxococcus xanthus
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(19019)
; OTHER INFORMATION: unsure at all n locations
US-09-902-540-1171

Alignment Scores:
Pred. No.: 2,29e-86 Length: 19019
Score: 728.00 Matches: 139
Percent Similarity: 77.49% Conservative: 40
Best Local Similarity: 60.17% Mismatches: 52
Query Match: 59.43% Indels: 0
DB: 4 Gaps: 0

US-10-035-300A-2 (1-239) x US-09-902-540-1171 (1-19019)
Qy 1 MetAlaThrProHisIleAsnAlaGluMetGlyAspPheAlaAspValValLeuMetPro 20
Db 16556 ATGGCGACTCTCTATATCTCGCTGCCCTGGTACTTCGCTGACGTGGTCTCTATCGCT 16497
Qy 21 GlyAspProLeuArgAlaLysTyrIleAlaGluThrPheLeuGluAspAlaArgGluVal 40
Db 16496 GGCACCGCTCCGCGCTCGTTACATCTCCGACCGCTTCTCGAAGGTGCCCGAGGTC 16437
Qy 41 AsnAsnValArgGlyMetLeuGlyPheThrGlyThrTyrLysGlyArgLysIleSerVal 60
Db 16436 ACCTCCGTGGCAACATGCTCGGCTTCCACGGGACCTTCCGGGGCCGGCGGTGTCGGTG 16377
Qy 61 MetGlyHisGlyValGlyIleProSerCysSerIleTyrThrLysGluLeuIleThrAsp 80
Db 16376 ATGGGGCAGCGCATGGGGTTCCTTCCATCTACGCCCGAGCTCATCAAGC 16317
Qy 81 PheGlyValLysLysIlelleArgValGlySerCysGlyAlaValLeuProHisValLys 100
Db 16316 TATGGGGTGGCGTGATCATCCGGTGGCGAGCTGCGCGCGCTGACGACCGCATGAAG 16257
Qy 101 LeuArgAspValValIleGlyMetGlyAlaCysThrAspSerLysValAsnArgIleArg 120
Db 16256 GTCCGGGAGGTGTCGTGCGACCGGGGGCGGCGGACGACTCCACGATGAATCGGATCGG 16197
Qy 121 PheLysAspHisAspPheAlaAlaIleAlaAspPheAspMetValArgAsnAlaValAsp 140
Db 16196 CTGATGGGCATGATCTTCGCGCGGTGGCGACTTCACGTGCGCGCGCGGCGCATGGAG 16137
Qy 141 AlaAlaLysAlaLeuGlyIleAspAlaArgValGlyAsnLeuPheSerAlaAspLeuPhe 160
Db 16136 GCGCGGAGGCGGCAACAGCTGTGCGCGCGCGCGCGCGCTTTCACGCTCCGACCTCTTC 16077
Qy 161 TyrSerProAspGlyGluMetPheAspValMetGluLysTyrGlyIleLeuGlyValGlu 180
Db 16076 TACCACCGCAGGAGGAGCTCAACGCCACCTTGGCGGGGATGGGTGTCCTGGCGCTCGAG 16017
Qy 181 MetGluAlaAlaGlyIleTyrGlyValAlaAlaGluPheGlyAlaLysAlaLeuThrIle 200
Db 16016 ATGGAGGTCCGCGGCTCTACGGGTGGCGCGAGTCCGCGCGCGCGCGCTGGGGCTG 15957
Qy 201 CysThrValSerAspHisIleArgThrHisGluGlnThrThrAlaAlaGluArgGlnThr 220
Db 15956 CTCACGGTGTGGACCATCATCACCAGGAGAGTCTCAGCGCGGAGAGTCTCAGCGCGGAG 15897
```

```
Qy 221 ThrPheAsnAspMetIleLysIleAlaLeuGlu 231
Db 15896 ACGTTCGACGAGATGATTGAATCGCCCTGGAC 15864
```

```
RESULT 9
US-09-107-532A-1185
; Sequence 1185, Application US/09107532A
; Patent No. 6583275
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 7310
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD/ROM ISO9660
; COMPUTER: PC
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,532A
; FILING DATE: 30-Jun-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/085,598
; FILING DATE: 14 May 1998
; APPLICATION NUMBER: 60/051571
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Ariniello, Pamela Deneka
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 1185:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 735 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Enterococcus faecium
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (B) LOCATION 1...735
; SEQUENCE DESCRIPTION: SEQ ID NO: 1185:
US-09-107-532A-1185

Alignment Scores:
Pred. No.: 3.39e-72 Length: 735
Score: 607.00 Matches: 122
Percent Similarity: 70.13% Conservative: 40
Best Local Similarity: 52.81% Mismatches: 67
Query Match: 49.55% Indels: 2
DB: 4 Gaps: 2

US-10-035-300A-2 (1-239) x US-09-107-532A-1185 (1-735)
Qy 5 HistLeAsnAlaGluMetGlyAspPheAlaAspValValLeuMetProGlyAspProLeu 24
Db 37 CATATGACGGAAGAGGCGGAGATCGCCGATAGATTTTACTTCGCGGAGATCCTTTG 96
Qy 25 ArgAlaLysTyrIleAlaGluThrPheLeuGluAspAlaArgGluValAsnValArg 44
```

Db 97 CGAGCTAATATATTCGCGAGACTTTTGTGAAGATCCAGTTTGTACATCAGTAGCG 156
Qy 45 GlyMetLeuGlyPheThrGlyThrTyrlsGlyArgLysGlyValMetGlyHisGly 64
Db 157 GGTATGTTAGGCTTTACCGCAATATATAAGGAGAACGTAATTCCTGTTCAAGGAACAGGA 216
Qy 65 ValGlyIleProSerCysSerIleTyrlsGlyLeuLeuIleThrAspPheGlyValys 84
Db 217 ATGGGAATGCCATCAGCCACCATCTATCCCATGAATTCATCCAGTCGTATGGTGAAG 276
Qy 85 LysIleIleArgValGlySerCysGlyValAlaValLeuProHisValLysLeuArgAspVal 104
Db 277 AAATGATTCAGTAGGACATCGCGTCATGTCMAAAGACGTCATGTCGTGAATTA 336
Qy 105 ValIleGlyMetGlyAlaCysThrAspSerLysValAsnArgIleArgPheLysAspHis 124
Db 337 GTACTAGCACAGGTGCAGCCAGCATCTTCGTGATGATCGAAGAAATTCGAAGCCTTT 396
Qy 125 AspPheAlaAlaIleAlaAspPheAspMetValArgAsnAlaValAlaAspAlaLysAla 144
Db 397 CATTTTCCTCCATCAGTGAATTCATCTATTAAGCAAGCATATGATAGCTAAAGAA 456
Qy 145 LeuGlyIleAspAlaArgValGlyAsnLeuPheSerAlaAspLeuPheTyrlsSerProasp 164
Db 457 AAGGCTATACGGTTCATGTAGGAATGTTCTTCGGAAGATTCCTCTCAAAAGATGAT 516
Qy 165 --GlyGluMetPheAspValMetGlyLysTyrlsGlyIleLeuGlyValGluMetGluAla 183
Db 517 TTGACAGAGAGCTTCCATTAAGTGA---CTTGGCGTTTAGGATGGAATGGAAGCA 573
Qy 184 AlaGlyIleTyrlsGlyValAlaAlaGluPheGlyAlaLysAlaLeuThrIleCysThrVal 203
Db 574 GCGGCTCTTATTAATCTAGGCGCAAAATACCATGTGCAGACATTAAGTTGATCAGCGTT 633
Qy 204 SerAspHisIleArgThrHisGluInThrThrAlaAlaGluArgGlnThrThrPheAsn 223
Db 634 AGCCACCATTTGATCACCAGGGAAGAAACACAGCTGCAGAACGTCATACGTTCAT 693
Qy 224 AspMetIleLysIleAlaLeuGluSerValLeu 234
Db 694 GAATGATCGAAGTAGGATTGGAAACACGCGATA 726

RESULT 10

US-09-134-000C-2608
; Sequence 2608, Application US/09134000C
; Patent No. 6617156
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; TITLE OF INVENTION: ENTEROCOCCUS FAECALIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 032796-032
; CURRENT APPLICATION NUMBER: US/09/134,000C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/055, 778
; PRIOR FILING DATE: 1997-08-15
; NUMBER OF SEQ ID NOS: 6812
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2608
; LENGTH: 714
; TYPE: DNA
; ORGANISM: Enterococcus faecalis
US-09-134-000C-2608

Alignment Scores:
Pred. No.: 4,71e-70 Length: 714
Score: 591.00 Matches: 117
Percent Similarity: 69.13% Conservative: 42
Best Local Similarity: 50.87% Mismatches: 71
Query Match: 48.24% Indels: 0
DB: Gaps: 0

US-10-035-300A-2 (1-239) x US-09-134-000C-2608 (1-714)

Qy 5 HisIleAsnAlaGluMetGlyAspPheAlaAspValValLeuMetProGlyAspProLeu 24
Db 16 CATATCGAAGCAAAACCCAGGTGAATTCGCCGATAAAATTTTACTACCAGGGGATCCCCCTC 75
Qy 25 ArgAlaLysTyrlsIleAlaGluThrPheLeuGluAspAlaArgGluValAsnAsnValArg 44
Db 76 CGAGCAAAATATATTCGCGAAACATTTTTAGAGAAATCTGTCTGCTATAATCAAGATTCGT 135
Qy 45 GlyMetLeuGlyPheThrGlyThrTyrlsGlyArgLysIleSerValMetGlyHisGly 64
Db 136 GGGATGTTAGGTTACACAGGACGATATAAGGACAGCGGTGTTTCAGTCCCAAGGAACAGGG 195
Qy 65 ValGlyIleProSerCysSerIleTyrlsGlyLeuLeuIleThrAspPheGlyValys 84
Db 196 ATGGGATGCCCTTCAGCAGGATTAATTCGCCCATGAGTTAATAACTCTTACGATGTGAAG 255
Qy 85 LysIleIleArgValGlySerCysGlyValAlaValLeuProHisValLysLeuArgAspVal 104
Db 256 AAGTTAATTCGTGTGAACGTTGTTCTATTTACAGAAAAGTCAACGTTTCGTGAATTA 315
Qy 105 ValIleGlyMetGlyAlaCysThrAspSerLysValAsnArgIleArgPheLysAspHis 124
Db 316 GTTATCGCACAGCAGCAGCGCGCTTTTCCCGGATTCGCAATGATTTTCTTAAATAT 375
Qy 125 AspPheAlaAlaIleAlaAspPheAspMetValArgAsnAlaValAlaAspAlaLysAla 144
Db 376 GATTTTCACAAATTCGAGCTTTGATCTTTTACTCAAACTTACGAAATTCCTTAAAGCA 435
Qy 145 LeuGlyIleAspAlaArgValGlyAsnLeuPheSerAlaAspLeuPheTyrlsSerProasp 164
Db 436 AAAGGCTTTTACTACCCATGTTGTAACGTTATGTCAGATGATGTTTTTACAAAGATAGT 495
Qy 165 GlyGluMetPheAspValMetGlyLysTyrlsGlyIleLeuGlyValGluMetGluAlaAla 184
Db 496 TTAGATGATGCTTTCGTTTAGTTCGATACGGGTATTTAGCGATGGAATCGAAGCGCA 555
Qy 185 GlyIleTyrlsGlyValAlaAlaGluPheGlyAlaLysAlaLeuThrIleCysThrValser 204
Db 556 ATGTTATATATTATAGCAGCAAAATTTGATGTAGAACATTTGGCGATTTATGACGCTGAGT 615
Qy 205 AspHisIleArgThrHisGluInThrThrAlaAlaGluArgGlnThrThrPheAsnAsp 224
Db 616 GATAGCTTGTGTCAGCGCAAGAAACACGCGCAGAAACGCCAAACGCGATTTTAAATGAA 675
Qy 225 MetIleLysIleAlaLeuGluSerValLeu 234
Db 676 ATGATTCGAGTTGTTTAGAAAACGGCAATT 705

RESULT 11

US-09-583-110-37
; Sequence 37, Application US/09583110
; Patent No. 669703
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al.
; TITLE OF INVENTION: Nucleic Acid and Amino Acid Sequences Relating to Streptococcus
; TITLE OF INVENTION: Pneumoniae for Diagnostics and Therapeutics
; FILE REFERENCE: PATH00-07A
; CURRENT APPLICATION NUMBER: US/09/583,110
; CURRENT FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/107,433
; PRIOR FILING DATE: 1998-06-30
; PRIOR APPLICATION NUMBER: US 60/085,131
; PRIOR FILING DATE: 1998-05-12
; PRIOR APPLICATION NUMBER: US 60/051,553
; PRIOR FILING DATE: 1997-07-02
; NUMBER OF SEQ ID NOS: 5322

SEQ ID NO 37
LENGTH: 711
TYPE: DNA
ORGANISM: Streptococcus pneumoniae
US-09-583-110-37

```
Alignment Scores:
Pred. No.: 4.26e-68 Length: 711
Score: 576.50 Matches: 115
Percent Similarity: 68.83% Conservative: 44
Best Local Similarity: 49.78% Mismatches: 71
Query Match: 47.06% Indels: 1
DB: 4 Gaps: 1

US-10-035-300A-2 (1-239) x US-09-583-110-37 (1-711)

Qy 5 HisileAsnAlaGluMetGlyAspPheAlaAspValValLeuMetProGlyAspProLeu 24
Db 10 CATATTGCTGCTCAGCAGCGTGAATTCGTGATAAAATCTTCTTCCTGGGATCCTCTT 69
Qy 25 ArgAlaLysTyrIleAlaGluThrPheLeuGluAspAlaArgGluValAsnValArg 44
Db 70 CGTGCTAAGTTTATTGCGGAGAAATTCCTTTGGTGATGCTGTTGTTTAAACGAAGTGGT 129
Qy 45 GlyMetLeuGlyPheThrGlyThrTyLysGlyArgLysIleSerValMetGlyHisGly 64
Db 130 AACATGTTTGGTTACACTGCTACTTACAGGGTCACCGTGTACTGTCTGCTGGAAGTGGG 189
Qy 65 ValGlyIleProSerCysSerIleTyrThrLysGluLeuIleThrAspPheGlyValLys 84
Db 190 ATGGGAATGCCATCTATTTCGATTATGCGGTGAGTTAATCGTAGACTACGGTGTGAAG 249
Qy 85 LysIleIleArgValGlySerCysGlyAlaValLeuProHisValLysLeuArgAspVal 104
Db 250 AAATTGATTCGTGTGGAACTGCAGGTTCTTTGAATGAAGAGGTTCATGTCGTGAATTA 309
Qy 105 ValIleGlyMetGlyAlaCysThrAspSerLysValAsnArgIleArgPheLysAspHis 124
Db 310 GTTTTGGCGCAGCGGTGCAACCACTCAACATCGTTGTAATGCTGCGCACAGTAC 369
Qy 125 AspPheAlaIleAlaAspPheAspMetValArgAsnAlaValAspAlaIleLysAla 144
Db 370 GATTTTCCACAAATGTGCTAGTTTGAATGATAAAGCTTACCATATCGCCMAAGAA 429
Qy 145 LeuGlyIleAspAlaArgValGlyAsnLeuPheSerLysAlaAspLeuPheTyrSerProAsp 164
Db 430 CTTGGTATGACTACTCAGTTGGGAACGTTTGTGTCATCTGCTGTTTACTCAAAATAC 489
Qy 165 GlyGlyMetPheAspValMetGluLysTyrGlyIleLeuGlyValGluMetGluAlaAla 184
Db 490 TTTGAAAGAATATCGAGCTTGTGAATGGGAGTCAAGGCTGTGGAATGGAGAGCA 549
Qy 185 GlyIleTyrGlyValAlaAlaGluPheGlyAlaLysAlaLeuThrIleCysThrValSer 204
Db 550 GCTCTTTACTATCTGTGCTGCCAATACCATGTTGATGCGCTAGCTATCATGACCATCTCT 609
Qy 205 AspHis---IleArgThrHisGluGlnThrThrAlaAlaGluArgGlnThrThrPheAsn 223
Db 610 GATAGCTTGTGCTCAATCCAGACGAGACACAACTGCAGAGAAGCGTCAAAATACCTTCACT 669
Qy 224 AspMetIleLysIleAlaLeuGluSerValLeu 234
Db 670 GATATGATGAAGGTGTGTTTGGAAACCTTGATT 702

RESULT 12
US-09-107-433-1967
; Sequence 1967, Application US/09107433
; Patent No. 6800744
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID
; SEQUENCES RELATING TO STREPTOCOCCUS PNEUMONIAE FOR DIAGNO
; THERAPEUTICS
; NUMBER OF SEQUENCES: 5206
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
```

Qy 125 AspPheAlaAlaIleAlaAaspPheAaspMetValArgAenAlaValAaspAlaAlaLysAla 144
Db 373 GATTTTCACAAATTCGTAGCTTTGATTTGCTGATTAAGCTACCATATCGCCAAAGAA 432
Qy 145 LeuGlyIleAaspAlaArgValGlyAenLeuPheSerAlaAaspLeuPheTySerProAasp 164
Db 433 CTTGGTATGACTACTCAGCTTGGGAACGTTTGTCTGATCTGAIGTCTTTTACTCAAAATAC 492
Qy 165 GlyGluMetPheAaspValMetGluLysTyrglyIleLeuGlyValGluMetGluAlaAla 184
Db 493 TTTGAAAGAAATATCGAGCTTGGTAAATGGGAGTCAAGGCTGTGGAAATGGAAGCAGCA 552
Qy 185 GlyIleTyrglyValAlaAlaGluPheGlyAlaLysAlaLeuThrIleCysThrValSer 204
Db 553 GCTCTTATCTATCTTGTGTCCTAATACCATCTGATGCGTAGCTATCATGACCATCTCT 612
Qy 205 AspHis---IleArgThrHisGluGlnThrThrAlaAlaGluArgGlnThrThrPheAen 223
Db 613 GATAGCTTGGTCAATCCAGACGAACACAACTGCAGAAAGAACGTCAAAATATACCTTCTACT 672
Qy 224 AspMetIleLysIleAlaLeuGluSerValLeu 234
Db 673 GATATGATGAAGGTGGTTCGAAACCTTGATT 705

RESULT 13

US-08-961-527-128
; Sequence 128, Application US/08961527
; Patent No. 6420135
; GENERAL INFORMATION:
; APPLICANT: Charles Kunsch
; TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 391
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/961.527
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 128:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 13440 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
US-08-961-527-128
Alignment Scores:
Pred. No.: 5,39e-66 Length: 13440
Score: 575.50 Matches: 115
Percent Similarity: 68.83% Conservative: 44
Best Local Similarity: 49.78% Mismatches: 71
Query Match: 46.98% Indels: 1
DB: 3 Gaps: 1

US-10-035-300A-2 (1-239) x US-08-961-527-128 (1-13440)
Qy 5 HisIleAenAlaGluMetGlyAaspPheAlaAaspValValLeuMetProGlyAaspProLeu 24
Db 10453 CATATTCCTGCTCAGCAGGTTGAATTCGTGATAAAATTCCTCTCTGGGGATCCCTCT 10512
Qy 25 ArgAlaLysTyrglyIleAlaGluThrPheLeuGluAaspAlaArgGluValAenAaspValArg 44
Db 10513 CGTGCTAAGTTTATTGCGGAGAAATTCCTTCATGATGCTGTTGTTTAAACGAAGTGGT 10572
Qy 45 GlyMetLeuGlyPheThrGlyThrTyrglyArgLysIleSerValMetGlyHisGly 64
Db 10573 AACATGTTGGTTACACTGCTACTTCAAGGGTCACTGTGTATCTGATCTGGAACCTGG 10632
Qy 65 ValGlyIleProSerCysSerIleTyrglyLeuLeuIleThrAaspPheGlyValLys 84
Db 10633 ATGGGAATGCATCTATTTGATTTATGCGGTGAGTTAATCTGTAGACTACGGTGTGAG 10692
Qy 85 LysIleIleArgValGlySerCysGlyAlaValLeuProHisValLysLeuArgAaspVal 104
Db 10693 AAATTGATTCGTGTGGAACTGCAGGTTCTTGAATGAAGAGGTTTCATGTTGTTGAATTA 10752
Qy 105 ValIleGlyMetGlyAlaCysThrAaspSerLysValAenArgIleArgPheLysAaspHis 124
Db 10753 GTTTTGGCGAGCGGCTGCAACCACTCAACATCGTTCGTAATGACTGGCCACAGTAC 10812
Qy 125 AspPheAlaAlaIleAlaAaspPheAaspMetValArgAenAlaValAaspAlaAlaLysAla 144
Db 10813 GATTTTCCAAATTTGTAGCTTTGATTTGCTGATAAAGCTACCATATGCGCCAAAATA 10872
Qy 145 LeuGlyIleAaspAlaArgValGlyAenLeuPheSerAlaAaspLeuPheTySerProAasp 164
Db 10873 CTTGGTATGACTACTCAGCTTGGGAACGTTTGTGTCATGTCATGCTTTTACTCAAATTAC 10932
Qy 165 GlyGluMetPheAaspValMetGluLysTyrglyIleLeuGlyValGluMetGluAlaAla 184
Db 10933 TTTGAAAGAAATATCGAGCTTGGTAAATGGGAGTCAAGGCTGTGGAATGGAAGCAGCA 10992
Qy 185 GlyIleTyrglyValAlaAlaGluPheGlyAlaLysAlaLeuThrIleCysThrValSer 204
Db 10993 GCTCTTATCTATCTTGTGCTGCAATACCATGTTGATGCGGTAGCTATCATGACCATCTCT 11052
Qy 205 AspHis---IleArgThrHisGluGlnThrThrAlaAlaGluArgGlnThrThrPheAen 223
Db 11053 GATAGCTTGGTCAATCCAGACGAACACAACTGCAGAAAGAACGTCAAAATATACCTTCACT 11112
Qy 224 AspMetIleLysIleAlaLeuGluSerValLeu 234
Db 11113 GATATGATGAAGGTGGTTCGAAACCTTGATT 11145

RESULT 14

US-09-134-001C-1402
; Sequence 1402, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 1402
; LENGTH: 732
; TYPE: DNA
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-1402

```
Alignment Scores:
Pred. No.: 1,81e-67 Length: 732
Score: 572.00 Matches: 123
Percent Similarity: 66.23% Conservative: 30
Best Local Similarity: 53.25% Mismatches: 76
Query Match: 46.69% Indels: 2
DB: 3 Gaps: 2

US-10-035-300A-2 (1-239) x US-09-134-001C-1402 (1-732)

Qy 2 AlaThrProHisIleAsnAlaGluMetGlyAspPheAlaAspValValLeuMetProGly 21
Db 31 AGTACACCTCACATTAAACCCAAATGGAGTTAAATAGCTAAATACTGTTAATGCCAGGC 90
Qy 22 AspProLeuArgAlaLysTyrIleAlaGluThrPheLeuGluAspAlaArgGluValAsn 41
Db 91 GATCCATTACGTCTCAATATATTGCTGAAATTTTTTAGAAAATGTTGAACAATTCAT 150
Qy 42 AsnValArgGlyMetLeuGlyPheThrGlyThrTyrLysGlyArgLysIleSerValMet 61
Db 151 ACAGTAGTAATATGTTGGTTACACAGGAACCTATTAAGGCCAAAGATTTCTGTGATG 210
Qy 62 GlyHisGlyValGlyIleProSerCysSerIleTyrThrLysGluLeuIleThrAspPhe 81
Db 211 GGCTCTGGGATGGGAATCCAAATGTTGTTATTTCTTATGAACTTTACCATTTCTTT 270
Qy 82 GlyValLysLysIleIleArgValGlySerCysGlyAlaValLeuProHisValLysLeu 101
Db 271 GATGTGGATACAAATCATTCGCGTAGGTTCATGTGGCGCTCTTCAAGAAGATGTCATTTA 330
Qy 102 ArgAspValValIleGlyMetGlyAlaCysThrAspLys---ValAsnArgIleArg 120
Db 331 TATGATGTGATTATGTCACAGCGCTCCAAACTCAATATATGTTGATCAATTTAAT 390
Qy 121 PheLysAspHisAspPheAlaIleAlaAspPheAspMetValArgAsnAlaValAsp 140
Db 391 ATACCGGGTCAC---TTTGGCGCTATTGCTGATTTAATCTAGTAGCTAAAGCTAAGAAA 447
Qy 141 AlaAlaLysAlaLeuGlyIleAspAlaArgValGlyAsnLeuPheSerAlaAspLeuPhe 160
Db 448 GCTGCTGATGAATTTGGTGCATATCATGATGAGGTAAACGTTTGTGTCATCTGATACAT 507
Qy 161 TyrSerProAspGlyGluMetPheAspValMetGluLysTyrGlyIleLeuGlyValGlu 180
Db 508 TATAATGCTGATCAACTTTTAATGATTCATGGAAAAGATGGGTATCTTAGGTATCGAA 567
Qy 181 MetGluAlaAlaGlyIleTyrGlyValAlaAlaGluPheGlyAlaLysAlaLeuThrIle 200
Db 568 ATGGAATCTGCAGGTCTATATTTAAATGCCATCCATGCCAATAAAAAGGCTTTAGGTATT 627
Qy 201 CysThrValSerAspHisIleArgThrHisGluGlnThrAlaAlaGluArgGlnThr 220
Db 628 TTTACAGTAAGTATCATATTTCTAAGAGATGAAGCCACTAGTCAGAAGAGACAAACA 687
Qy 221 ThrPheAsnAspMetIleLysIleAlaLeuGlu 231
Db 688 TCATTTACTCAATGATGGAATAGCGCTCGAA 720

RESULT 15
US-09-710-279-3511
; Sequence 3511, Application US/09710279
; Patent No. 6703492
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/09/710,279
; CURRENT FILING DATE: 2000-11-09
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3511

; LENGTH: 2962
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: nucleic acid sequence
US-09-710-279-3511

Alignment Scores:
Pred. No.: 3,96e-66 Length: 2962
Score: 569.00 Matches: 122
Percent Similarity: 66.23% Conservative: 31
Best Local Similarity: 52.81% Mismatches: 76
Query Match: 46.45% Indels: 2
DB: 4 Gaps: 2

US-10-035-300A-2 (1-239) x US-09-710-279-3511 (1-2962)

Qy 2 AlaThrProHisIleAsnAlaGluMetGlyAspPheAlaAspValValLeuMetProGly 21
Db 1638 AGTACACCTCACATTAAACCCAAATGGAGTTAAATAGCTAAATACTGTTAATGCCAGGC 1697
Qy 22 AspProLeuArgAlaLysTyrIleAlaGluThrPheLeuGluAspAlaArgGluValAsn 41
Db 1698 GACCCATTACGTCTCAATATATTGCTGAAATTTTTTAGAAAATGTTGAACAATTCAT 1757
Qy 42 AsnValArgGlyMetLeuGlyPheThrGlyThrTyrLysGlyArgLysIleSerValMet 61
Db 1758 ACAGTAGTAATATGTTGGTTACACAGGAACCTATTAAGGCCAAAGATTTCTGTGATG 1817
Qy 62 GlyHisGlyValGlyIleProSerCysSerIleTyrThrLysGluLeuIleThrAspPhe 81
Db 1818 GGCTCTGGGATGGGAATTCCAAGTATGTTGTTATTTCTTATGAACTTTACCATTTCTTT 1877
Qy 82 GlyValLysLysIleIleArgValGlySerCysGlyAlaValLeuProHisValLysLeu 101
Db 1878 GATGTGGATACAAATCATTCGCGTAGGTTCATGTGGCGCTCTTCAAGAAGATGTCATTTA 1937
Qy 102 ArgAspValValIleGlyMetGlyAlaCysThrAspLys---ValAsnArgIleArg 120
Db 1938 TATGATGTGATTATGTCACAGCGCTTCACAAACTCAAAATATGTTGATCAATTTAAT 1997
Qy 121 PheLysAspHisAspPheAlaIleAlaAspPheAspMetValArgAsnAlaValAsp 140
Db 1998 ATACCGGGTCAC---TTTGGCGCTATTGCTGATTTAATCTAGTAGCTAAAGCTAAGAAA 2054
Qy 141 AlaAlaLysAlaLeuGlyIleAspAlaArgValGlyAsnLeuPheSerAlaAspLeuPhe 160
Db 2055 GCTGCTGATGAATTTGGTGCATATCATGATGAGGTAAACGTTTGTGTCATCTGATACAT 2114
Qy 161 TyrSerProAspGlyGluMetPheAspValMetGluLysTyrGlyIleLeuGlyValGlu 180
Db 2115 TATAATGCTGATTCAACTTTTAATGATTCATGGAAGAAAGATGGGTATCTTAGGTATCGAA 2174
Qy 181 MetGluAlaAlaGlyIleTyrGlyValAlaAlaGluPheGlyAlaLysAlaLeuThrIle 200
Db 2175 ATGGAATCTGCAGGTCTATATTTAAATGCCATCCATGCCAATAAAAAGGCTTTAGGTATT 2234
Qy 201 CysThrValSerAspHisIleArgThrHisGluGlnThrAlaAlaGluArgGlnThr 220
Db 2235 TTTACAGTAAGTATCATATTTCTAAGAGATGAAGCCACTAGTCAGAAGAGACAAACA 2294
Qy 221 ThrPheAsnAspMetIleLysIleAlaLeuGlu 231
Db 2295 TCATTTACTCAATGATGGAATAGCGCTCGAA 2327

Search completed: July 23, 2005, 04:45:33
Job time : 1196 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: July 23, 2005, 03:20:48 ; Search time 742 Seconds
(without alignments)
2043.501 Million cell updates/sec
Title: US-10-035-300A-2
Perfect score: 1225
Sequence: 1 MATPHINAMGDPADVVLMP.....TTFNDMIKIALESVLLGXDE 239

Scoring table:
BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 7173243 seqs, 3172129809 residues

Total number of hits satisfying chosen parameters: 14346486

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool_p/US10035300/runat_18072005_165352_28093/app_query.fasta_1.391
-DB=Published Applications NA -QFMT=fastap -SUFFIX=rnpb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=Blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HRAPIZ=500 -MINLEN=0
-MAXLEN=200000000 -USER=US10035300@cgn1.1.723 @runat_18072005_165352_28093
-NCPU=6 -ICPU=3 -NO WMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPLOCK=100
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications NA:
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq.*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq.*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq.*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq.*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq.*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq.*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq.*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq.*
9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq.*
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq.*
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq.*
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq.*
13: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq.*
14: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq.*
15: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq.*
16: /cgn2_6/ptodata/2/pubpna/US10D_PUBCOMB.seq.*
17: /cgn2_6/ptodata/2/pubpna/US10E_PUBCOMB.seq.*
18: /cgn2_6/ptodata/2/pubpna/US10F_PUBCOMB.seq.*
19: /cgn2_6/ptodata/2/pubpna/US10G_PUBCOMB.seq.*
20: /cgn2_6/ptodata/2/pubpna/US10H_PUBCOMB.seq.*
21: /cgn2_6/ptodata/2/pubpna/US10I_PUBCOMB.seq.*
22: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq.*
23: /cgn2_6/ptodata/2/pubpna/US11A_PUBCOMB.seq.*
24: /cgn2_6/ptodata/2/pubpna/US11_NEW_PUB.seq.*
25: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq.*
26: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1225	100.0	720	17	US-10-035-300A-1
2	1222	99.8	720	21	US-10-035-300A-3
3	1222	99.8	3031	10	US-09-891-865A-12
4	1222	99.8	3031	24	US-11-036-497-12
5	1222	99.8	3128	10	US-09-891-865A-13
6	1222	99.8	3128	24	US-11-036-497-13
7	1222	99.8	3383	10	US-09-891-865A-3
8	1222	99.8	3383	24	US-11-036-497-3
9	1222	99.8	3934	10	US-09-891-865A-14
10	1222	99.8	3934	24	US-11-036-497-14
11	1222	99.8	4189	10	US-09-891-865A-5
12	1222	99.8	4189	24	US-11-036-497-5
13	1222	99.8	5013	14	US-10-198-034-5
14	1222	99.8	5013	15	US-10-198-228-5
15	1222	99.8	5241	10	US-09-891-865A-7
16	1222	99.8	5241	24	US-11-036-497-7
17	1222	99.8	5495	10	US-09-891-865A-4
18	1222	99.8	5495	24	US-11-036-497-4
19	1222	99.8	6046	10	US-09-891-865A-15
20	1222	99.8	6046	24	US-11-036-497-15
21	1222	99.8	6269	10	US-09-891-865A-9
22	1222	99.8	6269	24	US-11-036-497-9
23	1222	99.8	6299	10	US-09-891-865A-10
24	1222	99.8	6299	24	US-11-036-497-10
25	1222	99.8	6301	10	US-09-891-865A-6
26	1222	99.8	6301	24	US-11-036-497-6
27	1218	99.4	720	17	US-10-035-300A-3
28	1013	82.7	1830121	17	US-10-329-670-1
29	1013	82.7	1830121	20	US-10-158-863-1
30	1013	82.7	1830121	22	US-10-981-687-1
31	778	63.5	640681	9	US-09-730-988-1
32	700	57.1	5822	10	US-09-891-865A-8
33	700	57.1	5822	24	US-11-036-497-8
34	693.5	56.6	899	19	US-10-772-272A-15
35	675	55.1	843	9	US-09-895-913A-147
36	671.5	54.8	319630	17	US-10-398-221-7
37	671.5	54.8	3011208	17	US-10-398-221-2058
38	670.5	54.7	1171	17	US-10-398-221-3116
39	663	54.1	693	18	US-10-335-977-3426
40	601	49.1	16484	9	US-09-070-927A-158
41	593	48.4	687	18	US-10-335-977-3425
42	576.5	47.1	714	22	US-10-617-320-1967
43	575.5	47.0	708	21	US-10-472-928-1597
44	575.5	47.0	13440	8	US-08-961-527-128
45	575.5	47.0	13440	17	US-10-158-844-128

ALIGNMENTS

RESULT 1
US-10-035-300A-1
; Sequence 1, Application US/10035300A
; Publication No. US20030228576A1
; GENERAL INFORMATION:
; APPLICANT: Ealick, Steven E.
; APPLICANT: Parker, William B.
; APPLICANT: Secret, John A.
; APPLICANT: Sorscher, Eric J.
; TITLE OF INVENTION: MUTANT PURINE NUCLEOSIDE PHOSPHORYLASE PROTEINS AND CELLULAR DELIV
; FILE REFERENCE: UAB-20702/22
; CURRENT APPLICATION NUMBER: US/10/035.300A
; CURRENT FILING DATE: 2001-10-26
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn version 3.1

US-09-891-865A-12
; Sequence 12, Application US/09891865A
; Publication No. US20030059870A1
; GENERAL INFORMATION:
; APPLICANT: NORPHARMA SPA
; TITLE OF INVENTION: Recombinant bacterial strains for the production of
; TITLE OF INVENTION: natural nucleosides and modified analogues thereof
; FILE REFERENCE: 99DC26E
; CURRENT APPLICATION NUMBER: US/09/891.865A
; PRIOR FILING DATE: 2002-06-25
; PRIOR APPLICATION NUMBER: MI98A002792
; PRIOR FILING DATE: 1998-12-23
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn ver. 2.1
; SEQ ID NO 12
; LENGTH: 3031
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: udp and deoD
; OTHER INFORMATION: cloned into pGM746 without upstream ptac promoter
US-09-891-865A-12

Alignment Scores:
Pred. No.: 3.4e-142 Length: 3031
Score: 1222.00 Matches: 238
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 99.58% Mismatches: 0
Query Match: 99.76% Indels: 0
DB: 10 Gaps: 0

US-10-035-300A-2 (1-239) x US-09-891-865A-12 (1-3031)

Qy 1 MetAlaThrProHisIleAsnAlaGluMetGlyAspPheAlaAspValValLeuMetPro 20
Db 101 ATGGCTACCCACACATTAATGCAGAAATGGCGATTTCGTCGACGTAGTTTGTATGCCA 160

Qy 21 GlyAspProLeuArgAlaLysTyrIleAlaGluThrPheLeuGluAspAlaArgGluVal 40
Db 161 GCGACCCGCTGCGTGGCAAGTATATGCTGAACCTTTCCTTGAAGATGCCCGTGAAGTG 220

Qy 41 AsnAsnValArgGlyMetLeuGlyPheThrGlyThrLysGlyArgLysIleSerVal 60
Db 221 AACACGTTCCGCGTATGCTGGGCTTACCGGTACTTACAAAGCCGCAAAATTTCCGTA 280

Qy 61 MetGlyHisGlyValGlyIleProSerCysSerIleTyrThrLysGluLeuIleThrAsp 80
Db 281 ATGGCTACCGGTATGGGTATCCCGTCTGCTCCATCTACCAAGAACTGATCACCAGT 340

Qy 81 PheGlyValLysIleIleArgValGlySerCysGlyAlaValLeuProHisValLys 100
Db 341 TTCGCGTGAAGAAATATATCCGCGTGGGTTCCTTGAAGATGCCCGTGAAGTG 220

Qy 101 LeuArgAspValValIleGlyMetGlyPheThrGlyThrLysGlyArgLysIleSerVal 60
Db 401 CTGCGCACGCTGCTATCGGTATCGGTGCTGCGACCATTCCTTGAAGATGCCCGTGAAGTG 220

Qy 121 PheLysAspHisAspPheAlaAlaIleAlaAspPheAspMetValArgAsnAlaValAsp 140
Db 461 TTTAAGACCCATGACTTTGCGCTATCGCTGACTTCGACATGGTTCGCGTGAAGTG 220

Qy 141 AlaAlaLysAlaLeuGlyIleAspAlaArgValGlyAsnLeuPheSerAlaAspLeuPhe 160
Db 521 GCAGCTAAAGCAGCTGGGTATGATGCTCGCGTGGGTAACTGTTCTCGCGTGAAGTG 220

Qy 161 TyrSerProAspGlyGluMetPheAspValMetGluLysTyrGlyIleLeuGlyValGlu 180
Db 581 TACTCTCCGACGCGCAAAATGTTCCAGCTGATGAAATAATACCGCATTTCTCGCGTGAAGTG 220

Qy 181 MetGluAlaAlaGlyIleTyrGlyValAlaAlaGluPheGlyAlaLysAlaLeuThrIle 200
Db 641 ATGGAAAGCGCTGGTATCTACCGGCTGCTGCAGAAATTTGGCGCAAGAAAGCCCTGACCATC 700

Qy 201 CysThrValSerAspHisIleArgThrHisGluGlnThrThrAlaAlaGluArgGlnThr 220
Db 701 TGCACCGTATCTGACCATCCGACATCAGGAGCAGACCATCGCTGCGTGAAGTG 760

Qy 221 ThrPheAsnAspMetIleLysIleAlaLeuGluSerValLeuLeuGlyAspLysGlu 239
Db 761 ACCTTCAACGACATGATCAAAATCGCACTGGATCCGTTCTGCTGGGCGATAAAGAG 817

RESULT 4
US-11-036-497-12
; Sequence 12, Application US/11036497
; Publication No. US20050142645A1
; GENERAL INFORMATION:
; APPLICANT: Beatetti, Giuseppina
; APPLICANT: Cali', Simona
; APPLICANT: Ghisotti, Daniela
; APPLICANT: Orsini, Gaetano
; APPLICANT: Tonon, Giancarlo
; APPLICANT: Zuffi, Gabriele
; TITLE OF INVENTION: Recombinant bacterial strains for the production of natural nucle
; TITLE OF INVENTION: and modified analogues thereof
; FILE REFERENCE: 02901/000J410-USO
; CURRENT APPLICATION NUMBER: US/11/036.497
; CURRENT FILING DATE: 2005-01-14
; PRIOR APPLICATION NUMBER: PCT/EP99/10416
; PRIOR FILING DATE: 1999-12-23
; PRIOR APPLICATION NUMBER: MI98A002792
; PRIOR FILING DATE: 1998-12-23
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 12
; LENGTH: 3031
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: udp and deoD cloned into pGM746 without upstream ptac promoter
US-11-036-497-12

Alignment Scores:
Pred. No.: 3.4e-142 Length: 3031
Score: 1222.00 Matches: 238
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 99.58% Mismatches: 0
Query Match: 99.76% Indels: 0
DB: 24 Gaps: 0

US-10-035-300A-2 (1-239) x US-11-036-497-12 (1-3031)

Qy 1 MetAlaThrProHisIleAsnAlaGluMetGlyAspPheAlaAspValValLeuMetPro 20
Db 101 ATGGCTACCCACACATTAATGCAGAAATGGCGATTTCGTCGACGTAGTTTGTATGCCA 160

Qy 21 GlyAspProLeuArgAlaLysTyrIleAlaGluThrPheLeuGluAspAlaArgGluVal 40
Db 161 GCGACCCGCTGCGTGGCAAGTATATGCTGAACCTTTCCTTGAAGATGCCCGTGAAGTG 220

Qy 41 AsnAsnValArgGlyMetLeuGlyPheThrGlyThrLysGlyArgLysIleSerVal 60
Db 221 AACACGTTCCGCGTATGCTGGGCTTACCGGTACTTACAAAGCCGCAAAATTTCCGTA 280

Qy 61 MetGlyHisGlyValGlyIleProSerCysSerIleTyrThrLysGluLeuIleThrAsp 80
Db 281 ATGGCTACCGGTATGGGTATCCCGTCTGCTCCATCTACCAAGAACTGATCACCAGT 340

Qy 81 PheGlyValLysIleIleArgValGlySerCysGlyAlaValLeuProHisValLys 100
Db 341 TTCGCGTGAAGAAATATATCCGCGTGGGTTCCTTGAAGATGCCCGTGAAGTG 220

Qy 101 LeuArgAspValValIleGlyMetGlyAlaCysThrAspSerLysValAsnArgIleArg 120
Db 401 CTGCGCACGCTGCTATCGGTATGGGTGCTGCGACCATTCCTTGAAGATGCCCGTGAAGTG 220

Qy 121 PheLysAspHisAspPheAlaAlaIleAlaAspPheAspMetValArgAsnAlaValAsp 140
Db 461 TTTAAGACCCATGACTTTGCGCTATCGCTGACTTCGACATGGTTCGCGTGAAGTG 220

Qy 141 AlaAlaLysAlaLeuGlyIleAspAlaArgValGlyAsnLeuPheSerAlaAspLeuPhe 160
Db 521 GCAGCTAAAGCAGCTGGGTATGATGCTCGCGTGGGTAACTGTTCTCGCGTGAAGTG 220

Qy 161 TyrSerProAspGlyGluMetPheAspValMetGluLysTyrGlyIleLeuGlyValGlu 180
Db 581 TACTCTCCGACGCGCAAAATGTTCCAGCTGATGAAATAATACCGCATTTCTCGCGTGAAGTG 220

Qy 181 MetGluAlaAlaGlyIleTyrGlyValAlaAlaGluPheGlyAlaLysAlaLeuThrIle 200
Db 641 ATGGAAAGCGCTGGTATCTACCGGCTGCTGCAGAAATTTGGCGCAAGAAAGCCCTGACCATC 700

Db 461 TTTAAAGACCATGACTTTGGCGGTATCGCTGACTTCGACATCGTGGCTAAGCAGTAGAT 520
Qy 141 AlaAlaValAlaLeuGlyIleAspAlaAArgValGlyAsnLeuPheSerAlaAspLeuPhe 160
Db 521 GCAGCTAAAGCAGCTGGGTATGTGATGCTCGCGTGGGTAACCTGTTCCTCCGCTGACCTGTTT 580
Qy 161 TyrSerProAspGlyGluMetPheAspValMetGluLysTyrGlyIleLeuGlyValGlu 180
Db 581 TACTCTCCGACCGCGGAAATGTCGAGCTGATGGGAAATATACGGCATTCCTCGCGTGGAA 640
Qy 181 MetGluAlaAlaGlyIleTyrGlyValAlaAlaGluPheGlyAlaLysAlaLeuThrIle 200
Db 641 ATGGAAGCGCTGGTATCTACCGCGTCTGCGAGATTTGGCGGAAAGCCCTGACCATC 700
Qy 201 CysThrValSerAspHisIleAArgThrHisGluGlnThrThrAlaAlaGluArgGlnThr 220
Db 701 TGCACCGTATCTGACCATCCGACCTCACGAGCAGACCATCCCGCTGAGCGTCAGACT 760
Qy 221 ThrPheAsnAspMetIleLysIleAlaLeuGluSerValLeuLeuGlyAspLysGlu 239
Db 761 ACCTTCAACGACATGATCAAAATCGCACTGGAAATCGGTCTGCTGGCGGATAAAGAG 817

RESULT 5

US-09-891-865A-13
; Sequence 13, Application US/09891865A
; Publication No. US20030059870A1
; GENERAL INFORMATION:
; APPLICANT: NORPHARMA SPA
; TITLE OF INVENTION: Recombinant bacterial strains for the production of
; TITLE OF INVENTION: natural nucleosides and modified analogues thereof
; FILE REFERENCE: 99DC26E
; CURRENT APPLICATION NUMBER: US/09/891,865A
; CURRENT FILING DATE: 2002-06-25
; PRIOR APPLICATION NUMBER: MI98A002792
; PRIOR FILING DATE: 1998-12-23
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 13
; LENGTH: 3128
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: deoD cloned
; OTHER INFORMATION: downstream ptac promoter
US-09-891-865A-13

Alignment Scores:
Pred. No.: 3,55e-142 Length: 3128
Score: 1222.00 Matches: 238
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 99.58% Mismatches: 0
Query Match: 99.76% Indels: 0
DB: 10 Gaps: 0

US-10-035-300A-2 (1-239) x US-09-891-865A-13 (1-3128)

Qy 1 MetAlaThrProHisIleAsnAlaGluMetGlyAspPheAlaAspValValLeuMetPro 20
Db 198 ATGGCTACCCACACATTAATGAGAAATGGCGATTCGCTGACGTAGTTTGTATGCCA 257
Qy 21 GlyAspProLeuArgAlaLysTyrIleAlaGluThrPheLeuGluAspAlaArgGluVal 40
Db 258 GGGACCGCTGGTGGAGTATATGCTGAACCTTCCTTGAAGATGCCCTGAGTG 317
Qy 41 AsnAsnValArgGlyMetLeuGlyPheThrGlyThrTyrLysGlyArgLysIleSerVal 60
Db 318 AACACGCTTCGCGTATGCTGGGCTTCACCGTACTTACAAAGCGGCAAAATTCGCTA 377
Qy 61 MetGlyHisGlyValGlyIleProSerCysSerIleTyrThrLysGluLeuIleThrAsp 80
Db 378 ATGGGTACCGGTATGGGTATCCCGTCTGCTCCATCTACACCAAGAACTGATCCCGAT 437

Qy 81 PheGlyValLysLysIleIleArgValGlySerCysGlyAlaValLeuProHisValLys 100
Db 438 TTCGGCGTGAAGAAATTTATTCGGCGTGGGTTCCTGTGGCGCAGTTCTGCGCAGCTAAAA 497
Qy 101 LeuArgAspValValIleGlyMetGlyAlaCysThrAspSerLysValAsnArgIleArg 120
Db 498 CTGCGCGAGCTGCTTATTCGGTATGGTGGCTGCGCAGATTCCAAAGTTAACCGCATCCGT 557
Qy 121 PheLysAspHisAspPheAlaIleAlaAspPheAspMetValArgAsnAlaValAsp 140
Db 558 TTTAAAGACATGACTTTGCCGCTATCGCTGACTTCGACATGGTGGCTAACCGCAGTAGAT 617
Qy 141 AlaAlaLysAlaLeuGlyIleAspAlaAArgValGlyAsnLeuPheSerAlaAspLeuPhe 160
Db 618 GCAGCTAAAGCATGGGTATTCGCTCGCGTGGGTAACTGTTCCTCCGCTGACCTGTTT 677
Qy 161 TyrSerProAspGlyGluMetPheAspValMetGluLysTyrGlyIleLeuGlyValGlu 180
Db 678 TACTCTCCGACCGCGGAAATGTTCCGACGTGATGGAAAAATACGGCATTCCTCGCGTGGAA 737
Qy 181 MetGluAlaAlaGlyIleTyrGlyValAlaAlaGluPheGlyAlaLysAlaLeuThrIle 200
Db 738 ATGGAAGCGCTGGTATCTACGCGCTGCTGCAAAATTTGGCGCGAAAGCCCTGACCATC 797
Qy 201 CysThrValSerAspHisIleAArgThrHisGluGlnThrThrAlaAlaGluArgGlnThr 220
Db 798 TGCACCGTATCTGACCATCCGACCTCACGAGCAGACCATGCGCGCTGAGCGTCAGACT 857
Qy 221 ThrPheAsnAspMetIleLysIleAlaLeuGluSerValLeuLeuGlyAspLysGlu 239
Db 858 ACCTTCAACGACATGATCAAAATCGCACTGGAAATCGGTCTGCTGGCGCGATAAAGAG 914

RESULT 6

US-11-036-497-13
; Sequence 13, Application US/11036497
; Publication No. US20050142645A1
; GENERAL INFORMATION:
; APPLICANT: Bestetti, Giuseppe
; APPLICANT: Cali', Simona
; APPLICANT: Chisotti, Daniela
; APPLICANT: Orsini, Gaetano
; APPLICANT: Tonon, Giancarlo
; APPLICANT: Zuffi, Gabriele
; TITLE OF INVENTION: Recombinant bacterial strains for the production of natural nucle
; TITLE OF INVENTION: and modified analogues thereof
; FILE REFERENCE: 02901/000J410-US0
; CURRENT APPLICATION NUMBER: US/11/036,497
; CURRENT FILING DATE: 2005-01-14
; PRIOR APPLICATION NUMBER: PCT/EP99/10416
; PRIOR FILING DATE: 1999-12-23
; PRIOR APPLICATION NUMBER: MI98A002792
; PRIOR FILING DATE: 1998-12-23
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 13
; LENGTH: 3128
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: deoD cloned downstream ptac promoter
US-11-036-497-13

Alignment Scores:
Pred. No.: 3,55e-142 Length: 3128
Score: 1222.00 Matches: 238
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 99.58% Mismatches: 0
Query Match: 99.76% Indels: 0
DB: 24 Gaps: 0

US-10-035-300A-2 (1-239) x US-11-036-497-13 (1-3128)

Qy 1 MetAlaThrProHisIleAsnAlaGluMetGlyAspPheAlaAspValValLeuMetPro 20

```
Db 198 ATGCTACCCACACATTAAATGCGAAATGGCGATTTTCGCTACGAGTAGTTTTCATGCCA 257
Qy 21 GlyAspProLeuArgAlaLysTyrIleAlaGluThrPheLeuGluAspAlaArgGluVal 40
Db 258 GCGACCGCGTGGTGGAGTATATGCTGAACACTTCTTTGAAGATGCCGTTGAAGTG 317
Qy 41 AsnAsnValArgGlyMetLeuGlyPheThrGlyThrTyrLysGlyArgLysIleSerVal 60
Db 318 AACAACTTCGCGGTATGCTGGGCTTACCGGTATGCTTACAAAGCCGCAAAATTTCCGTA 377
Qy 61 MetGlyHisGlyValGlyIleProSerCysSerIleTyrThrLysGluLeuThrAsp 80
Db 378 ATGGGTACCGGTATGGGTATCCCGTCTGCTCCATCTACCAAGAACTGATCACCAGT 437
Qy 81 PheGlyValLysLysIleAlaArgValGlySerCysGlyAlaValLeuProHisValLys 100
Db 438 TTCGGCGTGAAGAAATATATCCGGTGGTTCCTGTGGCGAGTTCTCGCGCACGTAAGA 497
Qy 101 LeuArgAspValValIleGlyMetGlyAlaCysThrAspSerLysValAsnArgIleArg 120
Db 498 CTGGCGACGTCGTTATCGGTATGGTGGTCTGCACCGATTCCAAAGTTAACCGCATCCGT 557
Qy 121 PheLysAspHisAspPheAlaAlaIleAlaAspPheAspMetValArgAsnAlaValAsp 140
Db 558 TTTAAAGACCATGACTTTTGGCGCTATCGCTGACTTCGACATGGTGGCGTAACGAGTAGAT 617
Qy 141 AlaAlaLysAlaLeuGlyIleAspAlaArgValGlyAsnLeuPheSerAlaAspLeuPhe 160
Db 618 GCACCTAAAGCAGCTGGGTATGTGCTCGGTGGTAACTTCTCGCTGACCTGTTTC 677
Qy 161 TyrSerProAspGlyGluMetPheAspValMetGlyLysTyrGlyIleLeuGlyValGlu 180
Db 678 TACTCTCGGACGCGGAAATGTTTCGACGTATGATGCTCGGTGGTAACTTCTCGGCGTGGAA 737
Qy 181 MetGluAlaAlaGlyIleTyrGlyValAlaAlaGluPheGlyAlaLysAlaLeuThrIle 200
Db 738 ATGAAGCGGCTGTTATCTACGGCTCGCTGCAGAAATTTGGCGGCAAGCCCTGACCATC 797
Qy 201 CysThrValSerAspHisIleArgThrHisGluGlnThrThrAlaAlaGluArgGlnThr 220
Db 798 TGCACCGTATCTGACCATCTCCGACTCAGCAGCAGACCATCTGCGGTGAGCGTCAGACT 857
Qy 221 ThrPheAsnAspMetIleLysIleAlaLeuGluSerValLeuLeuGlyAspLysGlu 239
Db 858 ACCTTCAACGACATGATCAAAATCGCACTGGAATCCGTTCTGCTGGGCGATAAAGAG 914
```

RESULT 7

```
US-09-891-865A-3
; Sequence 3, Application US/09891865A
; Publication No. US20030059870A1
; GENERAL INFORMATION:
; APPLICANT: NORPHARMA SPA
; TITLE OF INVENTION: Recombinant bacterial strains for the production of
; FILE REFERENCE: natural nucleosides and modified analogues thereof
; CURRENT APPLICATION NUMBER: 99DC26E
; CURRENT FILING DATE: 2002-06-25
; PRIOR APPLICATION NUMBER: MI98A002792
; PRIOR FILING DATE: 1998-12-23
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 3383
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Plasmid
; NAME/KEY: gene
; LOCATION: (231)..(960)
; OTHER INFORMATION: deod
US-09-891-865A-3
```

Alignment Scores:

```
Pred. No.: 3,96e-142 Length: 3383
Score: 1222.00 Matches: 238
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 99.58% Mismatches: 0
Query Match: 99.76% Indels: 0
DB: 10 Gaps: 0
```

US-10-035-300A-2 (1-239) x US-09-891-865A-3 (1-3383)

```
Qy 1 MetAlaThrProHisIleAsnAlaGluMetGlyAspPheAlaAspValValLeuMetPro 20
Db 240 ATGGCTACCCACACATTAAATGCGAAATGGCGATTTTCGCTACGAGTAGTTTTCATGCCA 299
Qy 21 GlyAspProLeuArgAlaLysTyrIleAlaGluThrPheLeuGluAspAlaArgGluVal 40
Db 300 GCGACCGCGTGGTGGAGTATATGCTGAACACTTTCCTTGAAGATGCCGTTGAAGTG 359
Qy 41 AsnAsnValArgGlyMetLeuGlyPheThrGlyThrTyrLysGlyArgLysIleSerVal 60
Db 360 AACAACTTCGCGGTATGCTGGGCTTACCGGTATGCTTACAAAGCCGCAAAATTTCCGTA 419
Qy 61 MetGlyHisGlyValGlyIleProSerCysSerIleTyrThrLysGluLeuThrAsp 80
Db 420 ATGGGTACCGGTATGGGTATCCCGTCTGCTCCATCTACCAAGAACTGATCACCAGT 479
Qy 81 PheGlyValLysLysIleAlaArgValGlySerCysGlyAlaValLeuProHisValLys 100
Db 480 TTCGGCGTGAAGAAATATATCCGGTGGTTCCTGTGGCGAGTTCTTCCGCGCACGTAAGA 539
Qy 101 LeuArgAspValValIleGlyMetGlyAlaCysThrAspSerLysValAsnArgIleArg 120
Db 540 CTGGCGACGTCGTTATCGGTATGGTGGTCTGCACCGATTCCAAAGTTAACCGCATCCGT 599
Qy 121 PheLysAspHisAspPheAlaAlaIleAlaAspPheAspMetValArgAsnAlaValAsp 140
Db 600 TTTAAAGACCATGACTTTTGGCGCTATCGCTGACTTCGACATGGTGGCGTAACGAGTAGAT 659
Qy 141 AlaAlaLysAlaLeuGlyIleAspAlaArgValGlyAsnLeuPheSerAlaAspLeuPhe 160
Db 660 GCACCTAAAGCAGCTGGGTATGTGCTCGGTGGTAACTTCTCGCTGACCTGTTTC 719
Qy 161 TyrSerProAspGlyGluMetPheAspValMetGlyLysTyrGlyIleLeuGlyValGlu 180
Db 720 TACTCTCGGACGCGGAAATGTTTCGACGTATGATGCAAAAAATACGCCATTCTCGGCGTGGAA 779
Qy 181 MetGluAlaAlaGlyIleTyrGlyValAlaAlaGluPheGlyAlaLysAlaLeuThrIle 200
Db 780 ATGAAGCGGCTGGTATCTACGGCTCGCTGCAGAAATTTGGCGGCAAGCCCTGACCATC 839
Qy 201 CysThrValSerAspHisIleArgThrHisGluGlnThrThrAlaAlaGluArgGlnThr 220
Db 840 TGCACCGTATCTGACCATCTCCGACTCAGCAGCAGACCATCTGCGGTGAGCGTCAGACT 899
Qy 221 ThrPheAsnAspMetIleLysIleAlaLeuGluSerValLeuLeuGlyAspLysGlu 239
Db 900 ACCTTCAACGACATGATCAAAATCGCACTGGAATCCGTTCTGCTGGGCGATAAAGAG 956
```

RESULT 8

```
US-11-036-497-3
; Sequence 3, Application US/11036497
; Publication No. US20050142645A1
; GENERAL INFORMATION:
; APPLICANT: Beertti, Giuseppe
; APPLICANT: Cali', Simona
; APPLICANT: Ghisotti, Daniela
; APPLICANT: Orsini, Gaetano
; APPLICANT: Tonon, Giancarlo
; APPLICANT: Zuffi, Gabriele
; TITLE OF INVENTION: Recombinant bacterial strains for the production of natural nucle
; TITLE OF INVENTION: and modified analogues thereof
; FILE REFERENCE: 02901/000J410-US0
```

; CURRENT APPLICATION NUMBER: US/11/036,497
; CURRENT FILING DATE: 2005-01-14
; PRIOR APPLICATION NUMBER: PCT/EP99/10416
; PRIOR FILING DATE: 1999-12-23
; PRIOR APPLICATION NUMBER: MI98A002792
; PRIOR FILING DATE: 1998-12-23
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 3383
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: plasmid
; FEATURE:
; NAME/KEY: gene
; LOCATION: (231)..(960)
; OTHER INFORMATION: deod
US-11-036-497-3

Alignment Scores:
Pred. No.: 3,96e-142 Length: 3383
Score: 1222.00 Matches: 238
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 99.58% Mismatches: 0
Query Match: 99.76% Indels: 0
DB: 24 Gaps: 0

US-10-035-300A-2 (1-239) x US-11-036-497-3 (1-3383)

QY 1 MetAlaThrProHisIleAenAlaGluMetGlyAspPheAlaAspValValLeuMetPro 20
Db 240 ATGGCTACCCACACATTATGCAGAAATGGCGATTTCGCTGACGTAGTTTGTATGCCA 239
QY 21 GlyAspProLeuArgAlaLysTyrIleAlaGluThrPheLeuGluAspAlaArgGluVal 40
Db 300 GGGGACCCGCTGGCTGCGAAGTATATTGCTGAAACTTTCCTTGAAGATGCCCGTGAAGTG 359
QY 41 AsnAsnValArgGlyMetLeuGlyPheThrGlyThrTyrIysGlyArgLysIleSerVal 60
Db 360 AACAGCTTCGGGTATGCTGGGCTTCACCGGTACTTACAAAGCGCGCAAAATTTCCGTA 419
QY 61 MetGlyHisGlyValGlyIleProSerCysSerIleTyrThrLysGluLeuIleThrAsp 80
Db 420 ATGGGTACAGGTATGGGTATCCGCTCTGCTCCATCTACACCAAGACGTGATCCGAT 479
QY 81 PheGlyValLysLysIleIleArgValGlySerCysGlyAlaValLeuProHisValLys 100
Db 480 TTCGGCGTGAAGAAATATATCCGCGTGGGTTCCTGTGGCGCAGTTCTGCGCACGTAAAA 539
QY 101 LeuArgAspValValIleGlyMetGlyAlaCysThrAspSerLysValAsnArgIleArg 120
Db 540 CTCGGCGACGTGTTATCGGTATGGGTGCTGACCGATTCCAAAGTTAACCCGATCCCGT 599
QY 121 PheLysAspHisAspPheAlaIleAlaAspPheAspMetValArgAsnAlaValAsp 140
Db 600 TTTAAAGACCATGACTTTGCGGTATCGCTGATCTTCGACATGTGGGTAGCAGTAGAT 659
QY 141 AlaAlaLysAlaLeuGlyIleAspAlaArgValGlyAsnLeuPheSerAlaAspLeuPhe 160
Db 660 GCAGCTAAAGCACTGGGTATTGATGCTCGCGTGGGTAAACCTGTCTCCGCTGACCTGTTTC 719
QY 161 TyrSerProAspGlyGluMetPheAspValMetGluLysTyrGlyIleLeuGlyValGlu 180
Db 720 TACTCTCCGACCGCGAAATGTTCCGACGTGATGGAAAAATACGGCATTCCTCGCGGTGAA 779
QY 181 MetGluAlaAlaGlyIleTyrGlyValAlaAlaGluPheGlyAlaLysAlaLeuThrIle 200
Db 780 ATGGAAGCGGTGATCTACCGCGTTCGCTGACAGATTGGCGCGAAAGCCCTGACCATC 839
QY 201 CysThrValSerAspHisIleArgThrHisGluGlnThrThrAlaAlaGluArgGlnThr 220
Db 840 TGACCGGTATCTGACCACATCCGCACTCACGACAGACCACTGCCGCTGAGCGTCAGACT 899

QY 221 ThrPheAsnAspMetIleLysIleAlaLeuGluSerValLeuLeuGlyAspLysGlu 239
Db 900 ACCTTCAACGACATGATCAAAATCGCACTGGAAATCCGTTCTGCTGGCGGATAAAGAG 956

RESULT 9

US-09-891-865A-14
; Sequence 14, Application US/09891865A
; Publication No. US20030059870A1
; GENERAL INFORMATION:
; APPLICANT: NORPHARMA SPA
; TITLE OF INVENTION: Recombinant bacterial strains for the production of
; natural nucleosides and modified analogues thereof
; FILE REFERENCE: 99DC26E
; CURRENT APPLICATION NUMBER: US/09/891,865A
; CURRENT FILING DATE: 2002-06-25
; PRIOR APPLICATION NUMBER: MI98A002792
; PRIOR FILING DATE: 1998-12-23
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 14
; LENGTH: 3934
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: udp and deod
; OTHER INFORMATION: cloned downstream piac promoter
US-09-891-865A-14

Alignment Scores:

Pred. No.: 4,88e-142 Length: 3934
Score: 1222.00 Matches: 238
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 99.58% Mismatches: 0
Query Match: 99.76% Indels: 0
DB: 10 Gaps: 0

US-10-035-300A-2 (1-239) x US-09-891-865A-14 (1-3934)

QY 1 MetAlaThrProHisIleAenAlaGluMetGlyAspPheAlaAspValValLeuMetPro 20
Db 198 ATGGCTACCCACACATTATGCAGAAATGGCGATTTCGCTGACGTAGTTTGTATGCCA 257
QY 21 GlyAspProLeuArgAlaLysTyrIleAlaGluThrPheLeuGluAspAlaArgGluVal 40
Db 258 GGGGACCCGCTGGCTGCGAAGTATATTGCTGAACTTTCCTTGAAGATGCCCGTGAAGTG 317
QY 41 AsnAsnValArgGlyMetLeuGlyPheThrGlyThrTyrLysGlyArgLysIleSerVal 60
Db 318 AACCAAGCTTCGGGTATGCTGGGCTTCACCGGTACTTACAAAGCGCGCAAAATTTCCGTA 377
QY 61 MetGlyHisGlyValGlyIleProSerCysSerIleTyrThrLysGluLeuIleThrAsp 80
Db 378 ATGGGTACAGGTATGGGTATCCGCTCTGCTCCATCTACACCAAGACGTATCCGAT 437
QY 81 PheGlyValLysLysIleIleArgValGlySerCysGlyAlaValLeuProHisValLys 100
Db 438 TTCGGCGTGAAGAAATTTATCCGCGTGGGTTCCTGTGGCGCAGTTCTGCGCACGTAAAA 497
QY 101 LeuArgAspValValIleGlyMetGlyAlaCysThrAspSerLysValAsnArgIleArg 120
Db 498 CTCGGCGACGTGTTATCGGTATGGGTGCTGCGCGATTCCAAAGTTAACCGCATCCGT 557
QY 121 PheLysAspHisAspPheAlaIleAlaAspPheAspMetValArgAsnAlaValAsp 140
Db 558 TTTAAAGACCATGACTTTTCCGCTATCGCTGACTTCGACATGGTGGGTAGTACGCGATAGAT 617
QY 141 AlaAlaLysAlaLeuGlyIleAspAlaArgValGlyAsnLeuPheSerAlaAspLeuPhe 160
Db 618 GCAGCTAAAGCACTGGGTATTGATGCTCGCGTGGGTAACTGTCTCCGCTGACCTGTTTC 677
QY 161 TyrSerProAspGlyGluMetPheAspValMetGluLysTyrGlyIleLeuGlyValGlu 180

```
Db 678 TACTCTCCGACCGCGAAATGTTCCAGCTGATGGAAAAATACGGCAATTCCTCGGCGTGGA 737
Qy 181 MetGluAlaAlaGlyIleTyrglyValAlaAlaGluPheGlyAlaLysAlaLeuThrIle 200
Db 738 ATGGAAGCGGTGGTATCTACGGCTGCTGCAGAAATTTGGCGGAAAGCCCTGACCATC 797
Qy 201 CysThrValSerAspHisIleArgThrHisGluGlnThrThrAlaAlaGluArgGlnThr 220
Db 798 TGCACCGTATCTGACCACTCCGCACTCAGGACGAGACCACTGCGCGCTGAGCGTCA 857
Qy 221 ThrPheAsnAspMetIleLysIleAlaLeuGluSerValLeuLeuGlyAspLysGlu 239
Db 858 ACCTTCAACGACATGATCAAAATCGCACTGGAAATCCGTTCTGCTGGCGGATAAAG 914

RESULT 10
US-11-036-497-14
; Sequence 14, Application US/11036497
; Publication No. US20050142645A1
; GENERAL INFORMATION:
; APPLICANT: Bestetti, Giuseppe
; APPLICANT: Cali', Simona
; APPLICANT: Ghisotti, Daniela
; APPLICANT: Orsini, Gaetano
; APPLICANT: Tonon, Giancarlo
; APPLICANT: Zuffi, Gabriele
; TITLE OF INVENTION: Recombinant bacterial strains for the production of natural nucle
; FILE REFERENCE: 02901/0003410-USO
; CURRENT APPLICATION NUMBER: US/11/036,497
; CURRENT FILING DATE: 2005-01-14
; PRIOR APPLICATION NUMBER: PCR/EP99/10416
; PRIOR FILING DATE: 1999-12-23
; PRIOR APPLICATION NUMBER: M198A002792
; PRIOR FILING DATE: 1998-12-23
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 14
; LENGTH: 3934
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: udp and deod cloned downstream ptac promoter
US-11-036-497-14

Alignment Scores:
Pred. No.: 4,88e-142 Length: 3934
Score: 1222.00 Matches: 238
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 99.58% Mismatches: 0
Query Match: 99.76% Indels: 0
DB: 24 Gaps: 0

US-10-035-300A-2 (1-239) x US-11-036-497-14 (1-3934)
Qy 1 MetAlaThrProHisIleAsnAlaGluMetGlyAspPheAlaAspValValLeuMetPro 20
Db 198 ATGGCTACCCACACATTAATGCAAAATGGCGGATTCGCTGACCTAGTTTGTATGCGCA 257
Qy 21 GlyAspProLeuArgAlaLysTyriIleAlaGlnThrPheLeuGluAspAlaArgGluVal 40
Db 258 GCGACCCGCTGCGTGGCGAAGTATATTCCTGAAATCTTCTTGAAGATGCCCGTGAAGTG 317
Qy 41 AsnAsnValArgGlyMetLeuGlyPheThrGlyThrTyriLysGlyArgLysIleSerVal 60
Db 318 AACACGTTCCGGTATGCTGGCTTTCACCGGTACTTACAAAGCGCGCAAAATTTCCGTA 377
Qy 61 MetGlyHisGlyValGlyIleProSerCysSerIleTyriLysGluLeuIleThrAsp 80
Db 378 ATGGGTCAACGATATGGGTATCCCGTCTGCTCCATCATACCAAGAACTGATCACCAGT 437
Qy 81 PheGlyValLysLysIleIleArgValGlySerCysGlyAlaValLeuProHisValLys 100
Db 438 TTCGGCGTGAAGAAATATTCGGGTGGGTTCCTGTGGCGGAGTTCTGCGCGACGTAAA 497
```

```
Qy 101 LeuArgAspValValIleGlyMetGlyAlaCysThrAspSerLysValAsnArgIleArg 120
Db 498 CTGCGCGAGCTGGTATTCGGTATGGGTGCTGCACCGATTCCAAAGTTTAAACCGCATCCGT 557
Qy 121 PheLysAspHisAspPheAlaAlaIleAlaAspPheAspMetValArgAsnAlaValAsp 140
Db 558 TTTAAAGACCATGACTTTTGGCGCTATCGCTGACTTCGACATGGTGGTGGTAAACGACGTAGAT 617
Qy 141 AlaAlaLysAlaLeuGlyIleAspAlaArgValGlyAsnLeuPheSerAlaAspLeuPhe 160
Db 618 GCAGCTAAAGCACTGGGTATTTGATGCTGCGTGGGTAACTGTTCTCCGCTGACCTGTTTC 677
Qy 161 TyrSerProAspGlyGluMetPheAspValMetGluLysTyriGlyIleLeuGlyValGlu 180
Db 678 TACTCTCCGACCGCGCAAAATGTTCCAGCTGATGGAAAAATACGGCAATTCCTCGGCGTGGA 737
Qy 181 MetGluAlaAlaGlyIleTyrglyValAlaAlaGluPheGlyAlaLysAlaLeuThrIle 200
Db 738 ATGGAAGCGGTGGTATCTACGGCTGCTGCAGAAATTTGGCGGAAAGCCCTGACCATC 797
Qy 201 CysThrValSerAspHisIleArgThrHisGluGlnThrThrAlaAlaGluArgGlnThr 220
Db 798 TGCACCGTATCTGACCACTCCGCACTCAGGACGAGACCACTGCGCGCTGAGCGTCA 857
Qy 221 ThrPheAsnAspMetIleLysIleAlaLeuGluSerValLeuLeuGlyAspLysGlu 239
Db 858 ACCTTCAACGACATGATCAAAATCGCACTGGAAATCCGTTCTGCTGGCGGATAAAG 914

RESULT 11
US-09-891-865A-5
; Sequence 5, Application US/09891865A
; Publication No. US20030059870A1
; GENERAL INFORMATION:
; APPLICANT: NORPARMA SPA
; TITLE OF INVENTION: Recombinant bacterial strains for the production of
; FILE REFERENCE: 99DC26E
; CURRENT APPLICATION NUMBER: US/09/891,865A
; CURRENT FILING DATE: 2002-06-25
; PRIOR APPLICATION NUMBER: M198A002792
; PRIOR FILING DATE: 1998-12-23
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 4189
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Plasmid
; NAME/KEY: gene
; LOCATION: (243)..(1021)
; OTHER INFORMATION: udp
; NAME/KEY: gene
; LOCATION: (1037)..(1766)
; OTHER INFORMATION: deod
US-09-891-865A-5

Alignment Scores:
Pred. No.: 5,32e-142 Length: 4189
Score: 1222.00 Matches: 238
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 99.58% Mismatches: 0
Query Match: 99.76% Indels: 0
DB: 10 Gaps: 0

US-10-035-300A-2 (1-239) x US-09-891-865A-5 (1-4189)
Qy 1 MetAlaThrProHisIleAsnAlaGluMetGlyAspPheAlaAspValValLeuMetPro 20
Db 1046 ATGGCTACCCACACATTAATGCAAAATGGCGGATTCGCTGACCTAGTTTGTATGCGCA 1105
```


Qy 121 PheLysAspHisAspPheAlaAlaIleAlaAlaAspPheAspMetValArgAsnAlaValAsp 140
Db 773 TTTTAAAGACCATGACTTTGGCGCTATCGCTGACTTCGACATGGTGCGTAACGAGTAGAT 832
Qy 141 AlaAlaLysAlaLeuGlyIleAspAlaArgValGlyAsnLeuPheSerAlaAspLeuPhe 160
Db 833 GCAGCTAAAGCACCTGGGTATTGATGCTCGGTGGGTAAACCTGTTCTCCGCTGACCTGTTTC 892
Qy 161 TyrSerProAspGlyGluMetPheAspValMetGluLysTyrGlyIleLeuGlyValGlu 180
Db 893 TACTCTCCGACCGCGGAATGTTCCGAGCTGATGGAATAATACGCCATVTTCCGGCGTGAA 952
Qy 181 MetGluAlaAlaGlyIleTyrGlyValAlaAlaGluPheGlyAlaLysAlaLeuThrIle 200
Db 953 ATGGAACGGCTGTGATCTACGGCGTCTGCGAGATTTGGCGGAAGCCCTGACCATC 1012
Qy 201 CysThrValSerAspHisIleArgThrHisGluGlnThrThrAlaAlaGluArgGlnThr 220
Db 1013 TGCACCGTATCTGACCATCCGACCTCACGAGCAGACCACTGCGCTGAGCGTCCAGACT 1072
Qy 221 ThrPheAsnAspMetIleLysIleAlaLeuGluSerValLeuLeuGlyAspLysGlu 239
Db 1073 ACCTTCAACGACATGATCAAAATCGCACTCGAATCCGTTCTGCTGGCGGATAAAGAG 1129

RESULT 15

US-09-891-865A-7
; Sequence 7, Application US/09891865A
; Publication No. US20030059870A1
; GENERAL INFORMATION:
; APPLICANT: NORPHARMA SPA
; TITLE OF INVENTION: Recombinant bacterial strains for the production of
; TITLE OF INVENTION: natural nucleosides and modified analogues thereof
; FILE REFERENCE: 99DC26E
; CURRENT APPLICATION NUMBER: US/09/891.865A
; PRIOR FILING DATE: 2002-06-25
; PRIOR APPLICATION NUMBER: M198A002792
; PRIOR FILING DATE: 1998-12-23
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 7
; LENGTH: 5241
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Plasmid
; FEATURE:
; NAME/KEY: gene
; LOCATION: (1312)..(2042)
; OTHER INFORMATION: deod
US-09-891-865A-7

Alignment Scores:

Pred. No.:	7.26e-142	Length:	5241
Score:	1222.00	Matches:	238
Percent Similarity:	100.00%	Conservative:	1
Best Local Similarity:	99.58%	Mismatches:	0
Query Match:	99.76%	Indels:	0
DB:	10	Gaps:	0

US-10-035-300A-2 (1-239) x US-09-891-865A-7 (1-5241)

Qy 1 MetAlaThrProHisIleAsnAlaGluMetClyAspPheAlaAspValValLeuMetPro 20
Db 1322 ATGGCTATCCCCACACATTAATGCAAGAAATGGCGGATTTTCGCTGACGTAGTTTGTATGCCA 1381
Qy 21 GlyAspProLeuArgAlaLysTyrIleAlaGluThrPheLeuGluAspAlaArgGluVal 40
Db 1382 GGCAGCCCGCTGCGTGGGAAGTATATTGCTGAACATTTCTTCTGAAGATGCCCGTGAAGTG 1441
Qy 41 AsnAsnValArgGlyMetLeuGlyPheThrGlyThrTyrLysGlyArgLysIleSerVal 60
Db 1442 AACAAACGTTCCCGGTATGCTGGGCTTCACCGGTACTTACAAAGCGCGCAAAATTTCCGTA 1501

Qy 61 MetGlyHisGlyValGlyIleProSerCysSerIleTyrThrLysGluLeuIleThrAsp 80
Db 1502 ATGGGTACAGGTATGGGTATCCCGTCTGCTCTCATCTACACCAAGNACTGATCACCAGAT 1561
Qy 81 PheGlyValLysLysIleIleArgValGlySerCysGlyAlaValLeuProHisValLys 100
Db 1562 TTCGGCGTGAAGAAAATATTATCCGCGTGGGTTCCTGTGGCGCAGTTCTGCGCGCAGTAAAA 1621
Qy 101 LeuArgAspValValIleGlyMetGlyAlaCysThrAspSerLysValAsnArgIleArg 120
Db 1622 CTGCGCGACGTCGTATTATCGGTATGGGTGCTGACCCGATTTCCAAAGTTAACCCGATCCGT 1681
Qy 121 PheLysAspHisAspPheAlaAlaIleAlaAspPheAspMetValArgAsnAlaValAsp 140
Db 1682 TTTTAAAGACCATGACTTTGGCGCTATCGCTGACTTCGACATGGTGGCTAACCGAGTAGAT 1741
Qy 141 AlaAlaLysAlaLeuGlyIleAspAlaArgValGlyAsnLeuPheSerAlaAspLeuPhe 160
Db 1742 GCAGCTAAAGCACCTGGGTATTGATGCTCGCGTGGGTAAACCTGTTCTCCGCTGACCTGTTTC 1801
Qy 161 TyrSerProAspGlyGluMetPheAspValMetGluLysTyrGlyIleLeuGlyValGlu 180
Db 1802 TACTCTCCGACCGCGGAAAATGTTTCGACGTGATGGAATAATACGSCATTTCTCGCGGTGAA 1861
Qy 181 MetGluAlaAlaGlyIleTyrGlyValAlaAlaGluPheGlyAlaLysAlaLeuThrIle 200
Db 1862 ATGGAAGCGCTGCTGATCTACGGCGTTCGCTGCAAGATTTGGCGGAAAGCCCTGACCATC 1921
Qy 201 CysThrValSerAspHisIleArgThrHisGluGlnThrThrAlaAlaGluArgGlnThr 220
Db 1922 TGCACCGTATCTGACCACTCCGACCTCACGAGCAGACCACTGCGCGTGAAGCGTCAGACT 1981
Qy 221 ThrPheAsnAspMetIleLysIleAlaLeuGluSerValLeuLeuGlyAspLysGlu 239
Db 1982 ACCTTCAACGACATGATCAAAATCGCACTGGAATCCGTTCTGCTGGCGGATAAAGAG 2038

Search completed: July 23, 2005, 06:11:47

Job time : 757 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 18, 2005, 21:23:45 ; Search time 39 Seconds
(without alignments)
589.636 Million cell updates/sec

Title: US-10-035-300A-2
Perfect score: 1225
Sequence: 1 MATPHINAEMGDFADVVLMP.....TTFNDMIKIALESVLLGDKE 239
Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues
Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 79:*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1222	99.8	239	2	G91296
2	1222	99.8	239	2	A27854
3	1222	99.8	239	2	A86138
4	1187	96.9	239	2	AF1073
5	1106	90.3	239	2	AF0054
6	1085	88.6	239	2	JC4326
7	1017	83.0	238	2	B64074
8	993	81.1	241	2	C82087
9	858	70.0	245	2	C82505
10	778	63.5	234	2	A84993
11	693.5	56.6	233	2	D59614
12	675	55.1	233	2	B64667
13	675	55.1	305	2	D75306
14	671.5	54.8	233	2	AH1678
15	670.5	54.7	233	2	AH1306
16	663	54.1	233	2	C71848
17	659.5	53.8	234	2	JT0874
18	659	53.8	234	2	C86741
19	576.5	47.1	236	2	B97964
20	575.5	47.0	236	2	E95096
21	569	46.4	236	2	G90007
22	565.5	46.2	235	2	D89774
23	555.5	45.3	233	2	C71288
24	468	38.2	320	2	D64205
25	450	36.7	238	2	S73418
26	430.5	35.1	239	2	G82909
27	389	31.8	232	2	E90592
28	288.5	23.6	236	2	S58291
29	288	23.5	273	2	A84246

RESULT 1
G91296
purine-nucleoside phosphorylase [imported] - Escherichia coli (strain O157:H7, substrain C;Species: Escherichia coli
C;Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 09-Jul-2004
C;Accession: G91296
R;Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.; Gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A;Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genom A;Reference number: A99629; MUID:21156231; PMID:11258796
A;Accession: G91296
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-239 <HAY>
A;Cross-references: UNIPROT:P09743; GB:BA000007; PIDN:BA838766.1; PID:gl3364821; GSPDB:G A;Experimental source: strain O157:H7, substrain RIMD 0509952
C;Genetics:
A;Gene: ECs5343
C;Superfamily: purine-nucleoside phosphorylase pnp
Query Match 99.8%; Score 1222; DB 2; Length 239;
Best Local Similarity 99.6%; Pred. No. 4.7e-95;
Matches 238; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MATPHINAEMGDFADVVLMPGDPPLRAKYIAETFELEDAREVNNVRGMLGFTCTYKGRKISV 60
Db 1 MATPHINAEMGDFADVVLMPGDPPLRAKYIAETFELEDAREVNNVRGMLGFTCTYKGRKISV 60
Qy 61 MGHGVGIPSCSIYTKELITDFGVKKIIRVSGCGAVLPHVKLRDVVIGMGACTDSKVNRI 120
Db 61 MGHGVGIPSCSIYTKELITDFGVKKIIRVSGCGAVLPHVKLRDVVIGMGACTDSKVNRI 120
Qy 121 FKDHFAAIAADFDMVRNVAADKALGIDARVGNLFSADLFYSPDGMFDMVEKYGILGVE 180
Db 121 FKDHFAAIAADFDMVRNVAADKALGIDARVGNLFSADLFYSPDGMFDMVEKYGILGVE 180
Qy 181 MEANGIYGVAAEFGAKALTCTVSDHIRTHTHEQTAAERQTTFNDMIKIALESVLLGDKE 239
Db 181 MEANGIYGVAAEFGAKALTCTVSDHIRTHTHEQTAAERQTTFNDMIKIALESVLLGDKE 239

RESULT 2
A27854
purine-nucleoside phosphorylase (EC 2.4.2.1) - Escherichia coli (strain K-12)
N;Alternate names: inosine phosphorylase
C;Species: Escherichia coli
C;Date: 31-Mar-1988 #sequence_revision 03-Aug-1992 #text_change 09-Jul-2004
C;Accession: A41143; A27854; S56608; G65253
R;Hershfield, M.S.; Chaffee, S.; Koro-Johnson, L.; Mary, A.; Smith, A.A.; Short, S.A.
Proc. Natl. Acad. Sci. U.S.A. 88, 7185-7189, 1991
A;Title: Use of site-directed mutagenesis to enhance the epitope-shielding effect of cov

ALIGNMENTS

A;Reference number: A41143; MUID:91334430; PMID:1714590
A;Accession: A41143
A;Molecule type: DNA
A;Residues: 1-239 <HER>
A;Cross-references: UNIPROT:P09743; GB:M60917; NID:g147308; PIDN:AAA24401.1; PID:g147309
R;Larsen, J.E.L.; Albrechtsen, B.; Valentin-Hansen, P.
Nucleic Acids Res. 15, 5125-5140, 1987
A;Title: Analysis of the terminator region after the deoCABD operon of Escherichia coli
A;Reference number: A27854; MUID:87259990; PMID:3299264
A;Accession: A27854
A;Molecule type: DNA
A;Residues: 225-239 <LAR>
A;Cross-references: GB:X05629; NID:g41254; PIDN:CAA29114.1; PID:g41255
R;Burland, V.; Plunkett III, G.; Sofia, H.J.; Daniels, D.L.; Blattner, F.R.
Nucleic Acids Res. 23, 2105-2119, 1995
A;Title: Analysis of the Escherichia coli genome VI: DNA sequence of the region from 92.
A;Reference number: S56314; MUID:95334362; PMID:7610040
A;Accession: S56308
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-239 <BUR>
A;Cross-references: EMBL:U14003; NID:g1263172; PIDN:AAA97280.1; PID:g537224
A;Note: The nucleotide sequence was submitted to the EMBL Data Library, August 1994
R;Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Ferna, N.T.; Burland, V.; Riley, M.; Co
.A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A;Title: The complete genome sequence of Escherichia coli K-12.
A;Reference number: A64720; MUID:97426617; PMID:9278503
A;Accession: G65253
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-239 <BLAT>
A;Cross-references: GB:AE000508; GB:U00096; NID:g2367382; PIDN:NAC77337.1; PID:g1790844;
A;Experimental source: strain K-12, substrain MG1655
C;Genetics:
A;Gene: deoD
A;Map position: 100
A;Superfamily: purine-nucleoside phosphorylase pnp
C;Keywords: glycosyltransferase; pentosyltransferase; salvage pathway
Query Match 99.8%; Score 1222; DB 2; Length 239;
Best Local Similarity 99.6%; Pred. No. 4.7e-95;
Matches 238; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 MATPHINAEMGDFADVVLMPGDPPLRAKYIAETPLEDAREVNNVGMGLGFTGYGKRKISV 60
DB 1 MATPHINAEMGDFADVVLMPGDPPLRAKYIAETPLEDAREVNNVGMGLGFTGYGKRKISV 60
QY 61 MGHGVGIPSCSIYTKELITDFGVKKIIRVGCAGVLPVHVKLRDVIWGMGACTDSKVNRI 120
DB 61 MGHGMGIPSCSIYTKELITDFGVKKIIRVGCAGVLPVHVKLRDVIWGMGACTDSKVNRI 120
QY 121 FKDHDFAAIADFDMVRNAVDAAKALGIDARVGNLFSADLFYSPDGMFDMVMEKYIGLVE 180
DB 121 FKDHDFAAIADFDMVRNAVDAAKALGIDARVGNLFSADLFYSPDGMFDMVMEKYIGLVE 180
QY 181 MEAAGYGVAAEFAGAKALITCTVSDHIRTHTHEQTAAEROTTENDMIKIALESVLLGDKE 239
DB 181 MEAAGYGVAAEFAGAKALITCTVSDHIRTHTHEQTAAEROTTENDMIKIALESVLLGDKE 239
RESULT 3
A86138
purine-nucleoside phosphorylase [imported] - Escherichia coli (strain O157:H7, substrain
C;Species: Escherichia coli
C;Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 09-Jul-2004
C;Accession: A86138
R;Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
Nature 409, 529-533, 2001
A;Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A;Reference number: A85480; MUID:21074935; PMID:11206551
A;Accession: A86138

A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-239 <STO>
A;Cross-references: UNIPROT:P09743; GB:AE005174; NID:g12519411; PIDN:AAG59565.1; GSPDB:GN
A;Experimental source: strain O157:H7, substrain EDL933
C;Genetics:
A;Gene: deoD
A;Superfamily: purine-nucleoside phosphorylase pnp
Query Match 99.8%; Score 1222; DB 2; Length 239;
Best Local Similarity 99.6%; Pred. No. 4.7e-95;
Matches 238; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 MATPHINAEMGDFADVVLMPGDPPLRAKYIAETPLEDAREVNNVGMGLGFTGYGKRKISV 60
DB 1 MATPHINAEMGDFADVVLMPGDPPLRAKYIAETPLEDAREVNNVGMGLGFTGYGKRKISV 60
QY 61 MGHGVGIPSCSIYTKELITDFGVKKIIRVGCAGVLPVHVKLRDVIWGMGACTDSKVNRI 120
DB 61 MGHGMGIPSCSIYTKELITDFGVKKIIRVGCAGVLPVHVKLRDVIWGMGACTDSKVNRI 120
QY 121 FKDHDFAAIADFDMVRNAVDAAKALGIDARVGNLFSADLFYSPDGMFDMVMEKYIGLVE 180
DB 121 FKDHDFAAIADFDMVRNAVDAAKALGIDARVGNLFSADLFYSPDGMFDMVMEKYIGLVE 180
QY 181 MEAAGYGVAAEFAGAKALITCTVSDHIRTHTHEQTAAEROTTENDMIKIALESVLLGDKE 239
DB 181 MEAAGYGVAAEFAGAKALITCTVSDHIRTHTHEQTAAEROTTENDMIKIALESVLLGDKE 239
RESULT 4
AF1073
purine nucleoside phosphorylase (EC 2.4.2.1) [imported] - Salmonella enterica subsp. ent
C;Species: Salmonella enterica subsp. enterica serovar typhi
A;Note: this species has also been called Salmonella typhi
C;Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002
C;Accession: AF1073
R;Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher,
th, T.; Connerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,
.S.; Moule, S.; O'Gaora, P.
Nature 413, 848-852, 2001
A;Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;
A;Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serov
A;Reference number: AB0502; MUID:21534947; PMID:11677608
A;Accession: AF1073
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-239 <PAR>
A;Cross-references: GB:AL513382; PIDN:CAD03405.1; PID:g16505674; GSPDB:GN00176
C;Genetics:
A;Gene: deoD
A;Superfamily: purine-nucleoside phosphorylase pnp
C;Keywords: glycosyltransferase; pentosyltransferase
Query Match 96.9%; Score 1187; DB 2; Length 239;
Best Local Similarity 96.7%; Pred. No. 4e-92;
Matches 231; Conservative 4; Mismatches 4; Indels 0; Gaps 0;
QY 1 MATPHINAEMGDFADVVLMPGDPPLRAKYIAETPLEDAREVNNVGMGLGFTGYGKRKISV 60
DB 1 MATPHINAEMGDFADVVLMPGDPPLRAKYIAETPLEDAREVNNVGMGLGFTGYGKRKISV 60
QY 61 MGHGVGIPSCSIYTKELITDFGVKKIIRVGCAGVLPVHVKLRDVIWGMGACTDSKVNRI 120
DB 61 MGHGMGIPSCSIYTKELITDFGVKKIIRVGCAGVLPVHVKLRDVIWGMGACTDSKVNRI 120
QY 121 FKDHDFAAIADFDMVRNAVDAAKALGIDARVGNLFSADLFYSPDGMFDMVMEKYIGLVE 180
DB 121 FKDHDFAAIADFDMVRNAVDAAKALGIDARVGNLFSADLFYSPDGMFDMVMEKYIGLVE 180
QY 181 MEAAGYGVAAEFAGAKALITCTVSDHIRTHTHEQTAAEROTTENDMIKIALESVLLGDKE 239
DB 181 MEAAGYGVAAEFAGAKALITCTVSDHIRTHTHEQTAAEROTTENDMIKIALESVLLGDKE 239

```
RESULT 5
AF0054
Purine nucleoside phosphorylase (EC 2.4.2.1) [imported] - Yersinia pestis (strain CO92)
C:Species: Yersinia pestis
C>Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 09-Jul-2004
C:Accession: AF0054
R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B.;
deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.;
il, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrell,
Nature 413, 523-527, 2001
A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.
A:Reference number: AB0001; MUID:21470413; PMID:11586360
A:Accession: AF0054
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-239 <KUR>
A:Cross-references: UNIPROT:Q8ZIQ2; GB:AL590842; PIDN:CAC89297.1; PID:gl5978534; GSPDB:C
C:Genetics:
C:Gene: deoB
C:Superfamily: purine-nucleoside phosphorylase pnp
C:Keywords: glycosyltransferase; pentosyltransferase

Query Match 90.3%; Score 1106; DB 2; Length 239;
Best Local Similarity 90.7%; Pred. No. 2.5e-85;
Matches 215; Conservative 11; Mismatches 11; Indels 0; Gaps 0;

Qy 1 MATPHINAEMGDFADVLMFGDPLRAKYIAETFLFEDAREVNNVRGMLGFTGTYGKRSV 60
Dy 1 MATPHINAEMGDFADVLMFGDPLRAKYIAETFLFEDAREVNNVRGMLGFTGTYGKRSV 60

Qy 61 MGHGVIIPSCSIYTKELITDFGVKKIIRVSGCGAVLPHVKLRDVVIGMGACTDSKVNRI 120
Dy 61 MGHGVIIPSCSIYAKELITDFGVKKIIRVSGCGAVRTDVKLRDVVIGMGACTDSKVNRI 120

Qy 121 PKDHFAAIAADFMVRNAVDAAKALGIDARVGNLFSADLFYSPDGMFVMEKYGILGVE 180
Dy 121 PKDHFAAIAADFMVRNAVDAAKAGVNVVRGNLFSADLFYTPDPQNFVMEKYGILGVE 180

Qy 181 MEAGIYGVAAEFKAKALITCTVSDHIRTHTQTTAAERQTTFNDMIKIALESVLLGD 237
Dy 181 MEAGICGVAAEFKAKALITCTVSDHIRTHTQTTAAERQTTFNDMIKIALESVLLGD 237

RESULT 6
JC4326
Purine nucleoside phosphorylase (EC 2.4.2.1) - Klebsiella sp.
N:Alternate names: inosine phosphorylase
C:Species: Klebsiella sp.
C>Date: 06-Dec-1995 #sequence_revision 08-Feb-1996 #text_change 29-Sep-1999
C:Accession: JC4326; PC4103; S57413
R:Takehara, M.; Ling, F.; Iwawa, S.; Inoue, Y.; Kimura, A.
Biosci. Biotechnol. Biochem. 59, 1987-1990, 1995
A:Title: Molecular cloning and nucleotide sequence of purine nucleoside phosphorylase an
A:Reference number: JC4326; MUID:96068952; PMID:8534998
A:Accession: JC4326
A:Molecule type: DNA
A:Residues: 1-239 <TAK>
A:Cross-references: ENBL:X87881; NID:g871393; PIDN:CAA61136.1; PID:g871394
A:Accession: PC4103
A:Molecule type: protein
A:Residues: 1-10 <TA2>
C:Genetics:
C:Gene: pnp
C:Function:
C:Superfamily: purine-nucleoside phosphorylase pnp
C:Keywords: glycosyltransferase; pentosyltransferase

Query Match 88.6%; Score 1085; DB 2; Length 239;
Best Local Similarity 90.4%; Pred. No. 1.4e-83;
Matches 217; Conservative 7; Mismatches 14; Indels 2; Gaps 2;
```

```
Qy 1 MATPHINAEMGDFADVLMFGDPLRAKYIAETFLFEDAREVNNVRGMLGFTGTYGK 59
Dy 1 MATPHINAEMGDFADVLMFGDPLRANSIAETFLFEDAREVNNVRGMLGFTGTYGRRKIS 60

Qy 60 VMHGCVIPSCSIYTKELITDFGVKKIIRVSGCGAVLPHVKLRDVVIGMGACTDSKVNRI 119
Dy 61 VMHGCVIPSCSIYTKELITDFGVKKIIRVSGCGAVREDVKLRVWIGHGACTDSKVNRL 120

Qy 120 RFKDHDFAAATADFMVRNAVDAAKALGIDARVGNLFSADLFYSPDGMFVMEKYGILGV 179
Dy 121 RFKDHDFPAATADFMVRNAVD-AKALGVEARNSIFSADLFYTPDPSMFVMEKYGILGV 179

Qy 180 EMEAGIYGVAAEFKAKALITCTVSDHIRTHTQTTAAERQTTFNDMIKIALESVLLGDKE 239
Dy 180 EMEAGIYGVAAEFKAKALITCTVSDHIRTHTQTTAAERQTTFNDMIKIALESVLLDRKE 239

RESULT 7
B64074
Purine-nucleoside phosphorylase (EC 2.4.2.1) - Haemophilus influenzae (strain Rd KW20)
N:Alternate names: inosine phosphorylase
C:Species: Haemophilus influenzae
C>Date: 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 09-Jul-2004
C:Accession: B64074
R:Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage, A.;
Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman, J.;
D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghagen, N.S.M.
Science 269, 496-512, 1995
A:Authors: Gnehm, C.B.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter, C.
A:Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.
A:Reference number: A64000; MUID:95350630; PMID:7542800
A:Accession: B64074
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-238 <TIGR>
A:Cross-references: UNIPROT:P44417; GB:U32734; GB:I42023; NID:gl573498; PIDN:AAC23176.1;
C:Superfamily: purine-nucleoside phosphorylase pnp
C:Keywords: glycosyltransferase; pentosyltransferase; salvage pathway

Query Match 83.0%; Score 1017; DB 2; Length 238;
Best Local Similarity 84.3%; Pred. No. 6.9e-78;
Matches 198; Conservative 15; Mismatches 22; Indels 0; Gaps 0;

Qy 3 TPHINAEMGDFADVLMFGDPLRAKYIAETFLFEDAREVNNVRGMLGFTGTYGKRKISVMG 62
Dy 2 TPHINAPEGAFAADVLMFGDPLRAKYIAETFLQDVVEVTNVNRMGLGFTGTYGKRKISIMG 61

Qy 63 HGVIPSCSIYTKELITDFGVKKIIRVSGCGAVLPHVKLRDVVIGMGACTDSKVNRI RPK 122
Dy 62 HGMGIPSCSIYAKELITEYGVKKIIRVSGCGTVMRMDVKVRDVIIGLGACTDSKVNRI RPK 121

Qy 123 DHDFAAATADFMVRNAVDAAKALGIDARVGNLFSADLFYSPDGMFVMEKYGILGVEME 182
Dy 122 DNDFAAATADFMQAQVQAQAKGVVRVGNLFSADLFYTPDVENFVMEKYGILGVEME 181

Qy 183 AAGIYGVAAEFKAKALITCTVSDHIRTHTQTTAAERQTTFNDMIKIALESVLLGD 237
Dy 182 AAGIYGVAAEFKAKALITCTVSDHIRTHTQTTAAERQTTFNDMIKIALESVLLGD 236
```

```
RESULT 8
C82087
Purine nucleoside phosphorylase VC2347 [imported] - Vibrio cholerae (strain N16961 serog;
C:Species: Vibrio cholerae
C>Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 09-Jul-2004
C:Accession: C82087
R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J.;
hardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers, P.
l, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
Nature 406, 477-483, 2000
A:Title: DNA sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
A:Reference number: A82035; MUID:20406833; PMID:10952301
```

A;Accession: C82087
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-241 <HEI>
A;Cross-references: UNIPROT:Q9KPM0; GB:AE004305; GB:AE003852; NID:g9656912; PIDN:AAF9549
A;Experimental source: serogroup O1; strain N16961; biotype E1 Tor
C;Genetics:
A;Gene: VC2347
A;Map position: 1
C;Superfamily: purine-nucleoside phosphorylase pnp

Query Match 81.1%; Score 993; DB 2; Length 241;
Best Local Similarity 79.8%; Pred. No. 7.2e-76;
Matches 190; Conservative 26; Mismatches 22; Indels 0; Gaps 0;

QY 1 MATPHINAEMGDFADVLMFGDPLRAKYIAETFLDAREVNNVRLMGFTGTGKGRKISV 60
Db 1 MATPHINAEMGDFADVLMFGDPLRAKYIAETFLDAREVNNVRLMGFTGTGKGRKISV 60

QY 61 MGHGVGIPSCSIYTKELITDFGVKKIIRVSCGAVLPHVKLRDVLVGMGACTDSKNRIR 120
Db 61 MGHGVGIPSCSIYTKELITDFGVKKIIRVSCGAVLPHVKLRDVLVGMGACTDSKNRIR 120

QY 121 FKDHFAAATADFMVRNVAADAAKALGIDARVGNLFSADLFYSPDGMFDMVMEYKILGVE 180
Db 121 FKDHFAAATADFMVRNVAADAAKALGIDARVGNLFSADLFYSPDGMFDMVMEYKILGVE 180

QY 181 MEAAGIYGVAAEFAGAKALITCTVSDHIRTHTQTAARQTTFNDMIKIALESVLGDK 238
Db 181 MEAAGIYGVAAEFAGAKALITCTVSDHIRTHTQTAARQTTFNDMIKIALESVLGDK 238

RESULT 9
C82505
purine nucleoside phosphorylase VCA0053 [imported] - Vibrio cholerae (strain N16961 sero
C;Species: Vibrio cholerae
C;Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 09-Jul-2004
A;Accession: C82505
R;Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J.;
Chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bae, S.; Qin, H.; Dragoi, I.; Sellers, P.
L. R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
Nature 406, 477-483, 2000
A;Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
A;Reference number: A82035; MUID:20406833; PMID:10952301
A;Accession: C82505
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-245 <HEI>
A;Cross-references: UNIPROT:Q9KNB2; GB:AE004349; GB:AE003853; NID:g9657434; PIDN:AAF9596
A;Experimental source: serogroup O1; strain N16961; biotype E1 Tor
C;Genetics:
A;Gene: VCA0053
A;Map position: 2
C;Superfamily: purine-nucleoside phosphorylase pnp

Query Match 70.0%; Score 858; DB 2; Length 245;
Best Local Similarity 69.2%; Pred. No. 1.5e-64;
Matches 182; Conservative 36; Mismatches 36; Indels 0; Gaps 0;

QY 1 MATPHINAEMGDFADVLMFGDPLRAKYIAETFLDAREVNNVRLMGFTGTGKGRKISV 60
Db 10 MATPHINAEMGDFADVLMFGDPLRAKYIAETFLDAREVNNVRLMGFTGTGKGRKISV 69

QY 61 MGHGVGIPSCSIYTKELITDFGVKKIIRVSCGAVLPHVKLRDVLVGMGACTDSKNRIR 120
Db 70 MGHGVGIPSCSIYTKELITDFGVKKIIRVSCGAVLPHVKLRDVLVGMGACTDSKNRIR 129

QY 121 FKDHFAAATADFMVRNVAADAAKALGIDARVGNLFSADLFYSPDGMFDMVMEYKILGVE 180
Db 130 FSGHFAAATADFMVRNVAADAAKALGIDARVGNLFSADLFYSPDGMFDMVMEYKILGVE 189

QY 181 MEAAGIYGVAAEFAGAKALITCTVSDHIRTHTQTAARQTTFNDMIKIALESVL 234
Db 181 MEAAGIYGVAAEFAGAKALITCTVSDHIRTHTQTAARQTTFNDMIKIALESVL 234

Db 190 MEAAGIYGVAAADGARGALTILTVDHILRGELSSSEDROKSFNDMMKVALETAI 243

RESULT 10
A84993
purine-nucleoside phosphorylase (EC 2.4.2.1) [imported] - Buchnera sp. (strain APS)
C;Species: Buchnera sp.
C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 23-Mar-2001
C;Accession: A84993
R;Shigenobu, S.; Watanabe, H.; Hattori, M.; Sakaki, Y.; Ishikawa, H.
Nature 407, 81-86, 2000
A;Title: Genome sequence of the endocellular bacterial symbiont of aphids Buchnera sp. A
A;Reference number: A84930; MUID:20445173; PMID:10993077
A;Accession: A84993
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-234 <STO>
A;Cross-references: GB:AP000398; GSPDB:GN00144
A;Experimental source: strain APS
C;Genetics:
A;Gene: deoD; BU541
C;Superfamily: purine-nucleoside phosphorylase pnp
C;Keywords: glycosyltransferase; pentosyltransferase

Query Match 63.5%; Score 778; DB 2; Length 234;
Best Local Similarity 61.5%; Pred. No. 7.1e-58;
Matches 144; Conservative 43; Mismatches 47; Indels 0; Gaps 0;

QY 1 MATPHINAEMGDFADVLMFGDPLRAKYIAETFLDAREVNNVRLMGFTGTGKGRKISV 60
Db 1 MSTPHINSKDDFSDIVLMFGDPLRAKYIAETFLDAREVNNVRLMGFTGTGKGRKISV 60

QY 61 MGHGVGIPSCSIYTKELITDFGVKKIIRVSCGAVLPHVKLRDVLVGMGACTDSKNRIR 120
Db 61 MSHGIGIPASLYTRELIIEFNFKIIRITGCAVRDDIKLRDIVISMGSTDSKNRIR 120

QY 121 FKDHFAAATADFMVRNVAADAAKALGIDARVGNLFSADLFYSPDGMFDMVMEYKILGVE 180
Db 121 FNDHFAAATADFMVRNVAADAAKALGIDARVGNLFSADLFYSPDGMFDMVMEYKILGVE 180

QY 181 MEAAGIYGVAAEFAGAKALITCTVSDHIRTHTQTAARQTTFNDMIKIALESVL 234
Db 181 METAGIYGVASELKVALSICITVSDHITNKFLSKERESSFNDMIELALESVL 234

RESULT 11
D69614
purine nucleoside phosphorylase deoD - Bacillus subtilis
C;Species: Bacillus subtilis
C;Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 09-Jul-2004
C;Accession: D69614
R;Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Bertero
C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Choi
A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.
Nature 390, 249-256, 1997
A;Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gallerc
iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holzapfel, S.; Hosono, S.; Hullo, M.F.;
Koster, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois,
A;Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mausel,
Y. M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle,
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadate, Y.; Sato, T.; Scanlon,
A;Authors: Schleich, S.; Schroeter, R.; Scoffone, P.; Sekiguchi, J.; Sekowska, A.; Seror,
teuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpsira, P.; Tognoni, A.; Totsu, V.; Uchiyama,
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K.
A;Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.
A;Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.
A;Reference number: A69580; MUID:98044033; PMID:9384377
A;Accession: D69614
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-233 <KUN>
A;Cross-references: UNIPROT:O34925; GB:299114; GB:AL009126; NID:g2634230; PIDN:CAB13854-J
A;Experimental source: strain 168

C;Genetics: A;Gene: deob C;Superfamily: purine-nucleoside phosphorylase pnp

Query Match 56.6%; Score 693.5; DB 2; Length 233;
Best Local Similarity 58.1%; Pred. No. 8.5e-51; Mismatches 54; Indels 1; Gaps 1;
Matches 133; Conservative 41

Qy 5 HINAEIMGDPADVVLMPGDPPLRAKYIAETFLDAREVNNVRGMLGFTGTGKGRKISVMGHG 64
Db 4 HIGAEKGQIADTVLLPGDPLRAKYIAETFLDAREVNNVRGMLGFTGTGKGRKISVMGHG 63

Qy 65 VGIPSCSIYTKELITDFGVKKIIRVSCGAVLPHVKLRDVGVMGACTDTSKVNIRFKDH 124
Db 64 MGVPISISIVNELLQSYDVQNLIRVSCGARIKDVKVRDVLAMTSSDQMSQNRVAFGSV 123

Qy 125 DPAIAIDPDMVRNAVDAKALGIDARVGNLFSADLFYSPDGMFDMVMEKYGILGVMEAA 184
Db 124 DPAICADPELLKNAYDAKGVPTVGSVFTAQDFTNDSQI-EKLUKYGVLVGVMETT 182

Qy 185 GIYGVAEFGAKALITCTVSDHIRTHTHEQTAAERQTTFNNDMIKIALESV 233
Db 183 ALYTLAAKHGRKALSILTVSDHVLTGERTTAERQTTTFHDMIEVALHSV 231

RESULT 12
B64667
purine-nucleoside phosphorylase - Helicobacter pylori (strain 26695)
C;Species: Helicobacter pylori
C;Date: 09-Aug-1997 #sequence_revision 09-Aug-1997 #text_change 09-Jul-2004
C;Accession: B64667
R;Tomb, J.F.; White, O.; Kerlavage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.D.; Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khalak, H.G.; Glodek, A.; McKenne son, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.; Fujii, C.; Bowman, C.; Watthey, L. Nature 388, 539-547, 1997
A;Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser, C. A;Title: The complete genome sequence of the gastric pathogen Helicobacter pylori.
A;Reference number: A64520; MUID:97394467; PMID:9252185
A;Accession: B64667
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-233 <TOM>
A;Cross-references: UNIPROT:P56463; GB:AE000623; GB:AE000511; NID:g2314327; PIDN:AA00822
C;Superfamily: purine-nucleoside phosphorylase pnp

Query Match 55.1%; Score 675; DB 2; Length 233;
Best Local Similarity 55.0%; Pred. No. 3e-49; Mismatches 61; Indels 0; Gaps 0;
Matches 126; Conservative 42

Qy 3 TPHINAEIMGDPADVVLMPGDPPLRAKYIAETFLDAREVNNVRGMLGFTGTGKGRKISVMG 62
Db 2 TPHINAKIGDFYPQCLCGDPLRVSYIAKKFLQDAKEITNVNRNMLGFGSKYKGRGISLMG 61

Qy 63 HGVIPSCSIYTKELITDFGVKKIIRVSCGAVLPHVKLRDVGVMGACTDTSKVNIRFK 122
Db 62 HGMGIASCTIYVTELIKTYQVKELIRITGCAISPKVGLKDIIMATGASTDKNRVRFL 121

Qy 123 DHDPAAIADPDMVRNAVDAKALGIDARVGNLFSADLFYSPDGMFDMVMEKYGILGVME 182
Db 122 NHDLSATPDPFELSRAVQTAKRLGIDLKGVNFSDFYFSTHAFDLMAKNHIAEME 181

Qy 183 AAGIYGVAEFGAKALITCTVSDHIRTHTHEQTAAERQTTFNNDMIKIALE 231
Db 182 AAGLYATAMELNALAKCLCSVDHLITKEALSPKERVESPDNMIILALE 230

RESULT 13
D75306
purine nucleoside phosphorylase - Deinococcus radiodurans (strain R1)
C;Species: Deinococcus radiodurans
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 09-Jul-2004
C;Accession: D75306
R;White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;

M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Ma S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999
A;Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A;Reference number: A75250; MUID:20036896; PMID:10567266
A;Accession: D75306
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-305 <WHI>
A;Cross-references: UNIPROT:Q9RSF8; GB:AE002050; GB:AE000513; NID:g6459965; PIDN:AAF11716
A;Experimental source: strain R1
C;Genetics:
A;Gene: DR2166
A;Map position: 1
C;Superfamily: purine-nucleoside phosphorylase pnp

Query Match 55.1%; Score 675; DB 2; Length 305;
Best Local Similarity 54.8%; Pred. No. 4.1e-49; Mismatches 126; Conservative 43; Indels 0; Gaps 0;
Matches 133; Conservative 36

Qy 5 HINAEIMGDPADVVLMPGDPPLRAKYIAETFLDAREVNNVRGMLGFTGTGKGRKISVMGHG 64
Db 70 HLNAEKQIAETVLLPGDPLRAHYIAETFLDAREVNNVRGMLGFTGTGKGRKISVMGHG 129

Qy 65 VGIPSCSIYTKELITDFGVKKIIRVSCGAVLPHVKLRDVGVMGACTDTSKVNIRFKDH 124
Db 130 MGIPSCMIYVNELTQDYGCKNLIRVGTAGSYQEGVHVRDLVLAQAQACTSDSNINRVFER 189

Qy 125 DPAIADPDMVRNAVDAKALGIDARVGNLFSADLFYSPDGMFDMVMEKYGILGVMEAA 184
Db 190 TFPIADPELLRAYQIAQGRGSAHVGNISSDTFNDFTEFORWAEPGLVAVEMEA 249

Qy 185 GIYGVAEFGAKALITCTVSDHIRTHTHEQTAAERQTTFNNDMIKIALESV 234
Db 250 GLYTLAAKYGVRAITILTISDHLVTHEVTSAREERQLTFNGMIEVALDAAL 299

RESULT 14
AH1678
purine nucleoside phosphorylase [imported] - Listeria innocua (strain Clip11262)
C;Species: Listeria innocua
C;Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 09-Jul-2004
C;Accession: AH1678
R;Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloecker, D.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Feihi, H.; D.; Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A;Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Mat ok, C.; Schluter, T.; Simoes, N.; Tierrez, A.; Vasquez-Boland, J.A.; Voss, H.; Wehland, A;Title: Comparative genomics of Listeria species.
A;Reference number: AB1077; MUID:21537279; PMID:11679669
A;Accession: AH1678
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-233 <GLA>
A;Cross-references: UNIPROT:Q92AF2; GB:AL592022; PIDN:CAC97200.1; PID:g16414471; GSPDB:G7
A;Experimental source: strain Clip11262
C;Genetics:
A;Gene: deob
C;Superfamily: purine-nucleoside phosphorylase pnp

Query Match 54.8%; Score 671.5; DB 2; Length 233;
Best Local Similarity 56.6%; Pred. No. 5.9e-49; Mismatches 133; Conservative 36; Indels 11; Gaps 2;
Matches 133; Conservative 36

Qy 5 HINAEIMGDPADVVLMPGDPPLRAKYIAETFLDAREVNNVRGMLGFTGTGKGRKISVMGHG 64
Db 4 HIEAKQGEIAETILLPGDPLRAKYIAETFLDAREVNNVRGMLGFTGTGKGRKISVMGHG 63

Qy 65 VGIPSCSIYTKELITDFGVKKIIRVSCGAVLPHVKLRDVGVMGACTDTSKVNIRFKDH 124
Db 64 MGIPSIYVNELIQSYDVQNLIRVGTAGSYQEGVHVRDLVLAQAQACTSDSNINRVFAGV 123

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: July 22, 2005, 23:48:48 ; Search time 3877 Seconds
(without alignments)
2346.495 Million cell updates/sec

Title: US-10-035-300A-2

Perfect score: 1225
Sequence: 1 MATPHINAEMGDFADVLLMP.....TTFNDMIKIALESVLLGDKE 239

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO_epool_p/US10035300/runat_18072005_165351_27984/app_query.fasta_1.391
-DB=EST -QFMT=fastp -SUFFIX=rst -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10035300@cgn 1 1 5180 @runat_18072005_165351_27984 -NCPU=6 -ICPU=3
-NO MMAP -LARGQUERY -NEG SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

EST:*
1: gb_est1.*
2: gb_est2.*
3: gb_hcc.*
4: gb_est3.*
5: gb_est4.*
6: gb_est5.*
7: gb_est6.*
8: gb_g881.*
9: gb_g882.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	683	55.8	771	7	CV219418 EST879128
C 2	683	55.8	794	7	CV219419 EST879129
C 3	681	55.6	767	7	CV205284 EST864994
C 4	681	55.6	769	7	CV204676 EST864386
C 5	681	55.6	792	7	CV204675 EST864385
C 6	681	55.6	794	7	CV205285 EST864995
C 7	668	54.5	721	7	CV205285 EST864995
C 8	591	48.2	1589	8	BH770690 LLMtag44
C 9	436.5	35.6	878	8	AZ533877 ENTBJ77TR

10	436	35.6	921	8	BH138974	BH138974	ENTOV94TF
C 11	429.5	35.1	863	8	AZ680406	AZ680406	ENTGM40TF
C 12	376	30.7	861	8	BH166036	BH166036	ENTSY60TF
C 13	364	29.7	864	8	AZ675191	AZ675191	ENTFV59TR
C 14	343.5	28.0	865	8	AZ669801	AZ669801	ENTLL39TF
C 15	314.5	25.7	882	8	AZ681446	AZ681446	ENTMK91TR
C 16	296	24.2	879	8	AZ533664	AZ533664	ENTFT92TF
C 17	285	23.3	169	7	CF306721	CF306721	HDAL--04-
C 18	285	23.3	169	7	CF306860	CF306860	HDAL--05-
C 19	260	21.2	904	8	AZ536366	AZ536366	ENTCK45TF
C 20	242	19.8	161	7	CF307418	CF307418	HDAL--06-
C 21	239	19.5	861	8	AZ669012	AZ669012	ENTKL76TF
C 22	229.5	18.7	852	8	AZ542090	AZ542090	ENTDY06TF
C 23	223	18.2	801	9	CC924665	CC924665	t078f14ba
C 24	212	17.3	626	8	BH023711	BH023711	U1-revers
C 25	212	17.3	637	8	BH023716	BH023716	U4-revers
C 26	207.5	16.9	1189	9	CG754454	CG754454	P049-4-E1
C 27	205	16.7	898	8	BH148646	BH148646	ENTPE80TR
C 28	198	16.2	690	4	BH162943	BH162943	EST565466
C 29	197	16.1	734	4	BH159186	BH159186	EST561709
C 30	196	16.0	709	4	BH160915	BH160915	EST563438
C 31	196	16.0	742	4	BH170790	BH170790	EST573313
C 32	195	15.9	635	2	BF297837	BF297837	056PBH03
C 33	195	15.9	645	4	BH170428	BH170428	EST572951
C 34	195	15.9	768	4	BH159582	BH159582	EST562105
C 35	191.5	15.6	715	9	AG266877	AG266877	Pinegold1
C 36	189	15.4	740	8	BH023715	BH023715	U4-forwar
C 37	186	15.2	648	2	BF298132	BF298132	060PBG02
C 38	185.5	15.1	737	8	BH023713	BH023713	U3-forwar
C 39	185	15.1	604	4	BH162201	BH162201	EST564724
C 40	182.5	14.9	890	8	AZ529376	AZ529376	ENTCV88TF
C 41	178	14.5	549	6	CD344467	CD344467	EtEstef04
C 42	178	14.5	557	6	CD343688	CD343688	EtEstef08
C 43	177	14.4	226	7	CV213810	CV213810	EST873520
C 44	177	14.4	250	7	CV213811	CV213811	EST873521
C 45	177	14.4	572	4	BH165171	BH165171	EST567694

ALIGNMENTS

RESULT 1
CV219418/c
LOCUS CV219418 771 bp mRNA linear EST 16-SEP-2004
DEFINITION EST879128 non-normalized T1 cDNA library Trichomonas vaginalis cDNA
clone TVTCX26 3' end, mRNA sequence.
ACCESSION CV219418 GI:52166398
VERSION CV219418.1
KEYWORDS EST.
SOURCE Trichomonas vaginalis
ORGANISM Trichomonas vaginalis
Eukaryota; Parabasalidea; Trichomonada; Trichomonadida;
Trichomonadidae; Trichomonadinae; Trichomonas.
REFERENCE 1 (bases 1 to 771)
AUTHORS Carlton,J.M., Dyall,S., Johnson,P.J. and Fraser,C.M.
TITLE The complete genome sequence of the sexually transmitted parasite Trichomonas vaginalis
JOURNAL Unpublished (2004)
COMMENT Other ESTs: EST879129
Contact: Jane Carlton
Parasite Genomics Group
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-530-9319
Fax: 301-838-0208
Email: carlton@tigr.org
Seq primer: polydT 24 base.
Location/Qualifiers
1. 771
/organism="Trichomonas vaginalis"
/mol_type="mRNA"
/strain="T1"
/db_xref="taxon:5722"
/clone="TVTCX26"

/clone lib="non-normalized T1 cDNA library"
/note=Tvector: lambda Triplex2; Site 1: SfIA; Site 2:
SfIB; T. vaginalis strain T1 library constructed from
cDNA, made in lambda Triplex2. Inserts cloned
unidirectionally in the SfIA and SfIB sites. Mass excision
of library produced inserts in pTriplex2 plasmid. Inserts
sequenced from both 5' and 3' ends using Triplex2
sequencing primer and polydT 24 bp primer respectively."

ORIGIN

Alignment Scores:
Pred. No.: 1.77e-75 Length: 771
Score: 683.00 Matches: 134
Percent Similarity: 75.21% Conservative: 42
Best Local Similarity: 57.26% Mismatches: 58
Query Match: 55.76% Indels: 0
DB: 7 Gaps: 0

US-10-035-300A-2 (1-239) x CV219418 (1-771)

Qy 1 MetAlaThrProHisIleAsnAlaGluMetGlyAspPheAlaAspValValLeuMetPro 20
Db 759 ATGCAACACCCCACTTCTGCAAGGTTGGCGATTTCGTGAGACAGTTCATCTGTC 700

Qy 21 GlyAspProLeuArgAlaIleTyrIleAlaGluThrPheLeuGluAspAlaArgGluVal 40
Db 699 GCGATCCATCCGTGCAAGCTCATTTCTGTGTAACACTATCTTGAGACGCTAAGCAGTC 640

Qy 41 AsnAsnValArgGlyMetLeuGlyPheThrGlyThrTyrIleGlyArgIleSerVal 60
Db 639 AACTCTGTCTGTGAATGCTTGGTTTACAGGCACATACAGGCGAAACCACTCTCTGTT 580

Qy 61 MetGlyHisGlyValGlyIleProSerCysSerIleTyrThrIleGlyGluLeuIleThrAsp 80
Db 579 ATGGGTCTATGCGATGGGTATCCCATTTCCATCTACCGTGAAGAACTCTACAAAGCTT 520

Qy 81 PheGlyValIleGlyIleArgValGlySerCysGlyAlaValLeuProHisValIle 100
Db 519 TACAAGGTTAAGACCATCATTCGTGTCGCACTTGGCGACAGTTGATCCAAATGTTTAC 460

Qy 101 LeuArgAspValValIleGlyMetGlyAlaCysThrAspSerIleValAsnArgIleArg 120
Db 459 GTCCGCGATGCTGCAATGTGTCAGCATCTGGCAGCATACAAATGTTAAACAGAAATGCGC 400

Qy 121 PheIleAspHisAspPheAlaIleAlaAspPheAspMetValArgAsnAlaValAsp 140
Db 399 CTTCCTGGCCACGATTTCCCGACTGACGCCAATCTGGAAGTCGTAGTCCTTAGTTGAA 340

Qy 141 AlaAlaIleValAlaGlyIleAspAlaArgValGlyAsnLeuPheSerAlaAspLeuPhe 160
Db 339 TCCGCCAAGGCTCTCAACATCCCAACACAGGTTGGAAAGGCCCTACTCAACAGATATCTTC 280

Qy 161 TyrSerProAspGlyGluMetPheAspValMetGluIleTyrGlyIleLeuGlyValGlu 180
Db 279 TACAGCAAGGAACAAAGGATTAATGAGCGCTTGTCTCAGTACCACCTTCATTTGCTGTGCA 220

Qy 181 MetGluAlaAlaGlyIleTyrGlyValAlaAlaGluPheGlyAlaIleLeuThrIle 200
Db 219 ATGGAATCTCTGTCTCCCTCCCAATGCTGTATTAACATATATGCGCAAGAGTGGCTGATC 160

Qy 201 CysThrValSerAspHisIleArgThrHisGluGlnThrThrAlaAlaGluArgGlnThr 220
Db 159 TGCACAGTTTCGGATCATCATCACACAGAGTCCGCTACACAGAAAGAGCGCCAGACA 100

Qy 221 ThrPheAsnAspMetIleIleAlaLeuGluSerValLeu 234
Db 99 TCATTCCAGAACATGATCAAGATCGCTCTTGAAGCTACATTG 58

RESULT 2
CV219419
LOCUS CV219419 794 bp mRNA linear EST 16-SEP-2004
DEFINITION EST879129 non-normalized T1 cDNA library Trichomonas vaginalis cDNA
clone TVTXX26 5' end, mRNA sequence.

CV219419
CV219419.1 GI:52166399
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
FEATURES
source
ORIGIN
Alignment Scores:
Pred. No.: 1.85e-75 Length: 794
Score: 683.00 Matches: 134
Percent Similarity: 75.21% Conservative: 42
Best Local Similarity: 57.26% Mismatches: 58
Query Match: 55.76% Indels: 0
DB: 7 Gaps: 0

US-10-035-300A-2 (1-239) x CV219419 (1-794)

Qy 1 MetAlaThrProHisIleAsnAlaGluMetGlyAspPheAlaAspValValLeuMetPro 20
Db 13 ATGCAACACCCCACTTCTGCAAGGTTGGCGATTTCGTGAGACAGTTCATCTGTC 72

Qy 21 GlyAspProLeuArgAlaIleTyrIleAlaGluThrPheLeuGluAspAlaArgGluVal 40
Db 73 GCGATCCATCCGTGCAAGCTCATTTCTGTGTAACACTATCTTGAGAACGCTAAGCAGTC 132

Qy 41 AsnAsnValArgGlyMetLeuGlyPheThrGlyThrTyrIleGlyArgIleSerVal 60
Db 133 AACTCTGTCTGTGAATGCTTGGTTTACAGGCACATACAGGCGAAACCACTCTCTGTT 192

Qy 61 MetGlyHisGlyValGlyIleProSerCysSerIleTyrThrIleGlyGluLeuIleThrAsp 80
Db 193 ATGGGTCTATGCGATGGGTATCCCATTTCTACCTGACGAGAACTCTCAACGTT 252

Qy 81 PheGlyValIleGlyIleArgValGlySerCysGlyAlaValLeuProHisValIle 100
Db 253 TACAAGGTTAAGACCATCATTCGTGTCGACATTCGCGCACAGTTGATCCAAATGTTTAC 312

Qy 101 LeuArgAspValValIleGlyMetGlyAlaCysThrAspSerIleValAsnArgIleArg 120
Db 313 GTCCGCGATGCTGCAATGTGTCAGCATCTGGCAGCATTCAAATGTTTACAGAAATGCGC 372

Qy 121 PheIleAspHisAspPheAlaIleAlaAspPheAspMetValArgAsnAlaValAsp 140

373	Db	CTTCTTGGCCACGATTTCCCGAGCTACAGCCAACTTCGAAGTCGTTAGTGCCCTTAGTTGAA	433
141	Qy	AlaAlaLysAlaLeuGlyIleAspAlaArgValGlyAsnLeuPheSerAlaAspLeuPhe	160
433	Db	TCCGCCAAGGCTCTCAACATCCCAACACAGGTTGGAAGGCCCTACTCAACAGATATCTTC	492
161	Qy	TyrSerProAspGlyGluMetPheAspValMetGluLysTyrGlyIleLeuGlyValGlu	180
493	Db	TACAGCAAGGAACCAAGGATTAATGAGGCCCTTGCTCAGTACCACCTTCATTGCTCTCGAA	552
181	Qy	MetGluAlaAlaGlyIleTyrGlyValAlaAlaGluPheGlyAlaLysAlaLeuThrIle	200
553	Db	ATGGAATCTGTGTGTCCTTCCCAATTGCTGATACTATGGCGCAAGAGCTGGCTGCATC	612
201	Qy	CysThrValSerAspHisIleArgThrHisGluGlnThrThrAlaAlaGluArgGlnThr	220
613	Db	TGCACAGTTTCGGATCATCATCATCACACACGAGTCCGCTACACCAGAGAGCGCCACACA	672
221	Qy	ThrPheAsnAspMetIleLysIleAlaLeuGluSerValLeu	234
673	Db	TCATTTCAGAACATGATCAAGATCGCTCTCAAGCTACATATG	714

RESULT 3	
CV205284/c	
LOCUS	
DEFINITION	CV205284 767 bp mRNA linear EST 16-SEP-2004
	EST864994 non-normalized T1 cDNA library Trichomonas vaginalis cDNA
	clone TTVAC60 3' end, mRNA sequence.

VERSION	KEYWORDS	SOURCE	ORGANISM
CV205284.1	GI:52152264	EST.	
		Trichomonas vaginalis	
		Trichomonas vaginalis	
		Eukaryota; Parabasalidea; Trichomonada; Trichomonadida; Trichomonadidae; Trichomonadinae; Trichomonas.	

REFERENCE	1 (bases 1 to 767)
AUTHORS	Carlton, J.M., Dyal, S., Johnson, P.J. and Fraser, C.M.
TITLE	The complete genome sequence of the sexually transmitted parasite <i>Trichomonas vaginalis</i>
JOURNAL	Unpublished (2004)
COMMENT	Other ESTs: EST864995

Contact: Jane Carlton
Parasite Genomics Group
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-530-9319
Fax: 301-838-0208
Email: carlton@tigr.org
Seg primer: polydT 24 base.

```

FEATURES
    source
        Location/Qualifiers
            1..767
                /organism="Trichomonas vaginalis"
                /mol_type="mRNA"
                /strain="T1"
                /db_xref="taxon:5722"
                /clone="TVTAC60"
                /clone_lib="non-normalized T1 cDNA library"
                /note="Vector: Lambda Triplex2; Site 1: SfiA; Site 2:
                SfiB; T. vaginalis strain T1 library constructed from
                cDNA, made in lambda Triplex2. Inserts cloned
                unidirectionally in the SfiA and SfiB sites. Mass excision
                of library produced inserts in pTriplex2 plasmid. Inserts
                sequenced from both 5' and 3' ends using Triplex2
                sequencing primer and polydT 24 bp primer respectively."

```

SEQUENCING PRIMER AND POSITIVELY AMPLIFIED DNA					
ORIGIN					
Alignment Scores:					
Pred. No.:	3.15e-75	Length:	767		
Score:	681.00	Matches:	134		
Percent Similarity:	75.21%	Conservative:	42		
Best Local Similarity:	57.26%	Mismatches:	58		
Query Match:	55.59%	Indels:	0		
DB:	7	Gaps:	0		

US-10-035-300A-2 (1-239) x CV205284 (1-767)

QY	1	MetAlaThrProHisIleAsnAlaGluMetGlyAspPheAlaAspValValLeuMetPro	20
DB	755	ATGCAACACCCCACTAATCTGCAAAAGTTGGCGATTTCGCTGAGACAGATTCTCATGTGC	696
QY	21	GlyAspProLeuArgAlaLysTyrIleAlaGluThrPheLeuGluAspAlaArgGluVal	40
DB	695	GGCGATCCACTCCGTGCAAGCTCATTTGATAACTATCTTTGAGAAGCGCTAAGCAAGTC	636
QY	41	AsnAsnValArgGlyMetLeuGlyPheThrGlyThrTyrLysGlyArgLysIleSerVal	60
DB	635	AATCTGTTCTGGATGCTTGGTTTACAGGCACATACAGAGGCAAAACCACCTCTCTGTT	576
QY	61	MetGlyHisGlyValGlyIleProSerCysSerIleTyrThrLysGluLeuIleThrAsp	80
DB	575	ATGGGTGTCATGGCATGGGTATCCCATCATTTCCATCTACGCTGAAGAACTCTTACAAACGTT	516
QY	81	PheGlyValLysLysIleIleArgValGlySerCysGlyAlaValLeuProHisValLys	100
DB	515	TACAGGTTAAGACCATCATCCGTGTCGGCACTTGGCGCACAGTTGATCCAAATGTTCCAC	456
QY	101	LeuArgAspValValIleGlyMetGlyAlaCysThrAspSerLysValAsnArgIleArg	120
DB	455	GTCCGGCATGCTCTGCATTGTCACAGCATCTGGCACAGATTTCAATGTTTAACAGAATGCGC	396
QY	121	PheLysAspHisAspPheAlaIleAlaAspPheAspMetValArgAsnAlaValAsp	140
DB	395	CTTCTTGGCCACAGATTTCCAGCTTACAGCCAACTTCGAAGTCGTAGTGCCCTTAGTTGAA	336
QY	141	AlaAlaLysValAlaLeuGlyIleAspAlaArgValGlyAsnLeuPheSerAlaAspLeuPhe	160
DB	335	TCCGCCAAGGCTCTCAACATCCCAACAGGCTTGGAAAGGCCTACTCAACAGATATCTTC	276
QY	161	TyrSerProAspGlyGluMetPheAspValMetGluLysTyrGlyIleLeuGlyValGlu	180
DB	275	TACAGCAAGGAAACAAAGGATTAATAGGCCCTTCTGCTCAGTACCACCTTCATTGCTGCGAA	216
QY	181	MetGluAlaAlaGlyIleTyrGlyValAlaAlaGluPheGlyAlaLysAlaLeuThrIle	200
DB	215	ATGGATCTGCTGGTCCCTTCCCAATTTGCTGATTTACTATGGCGCAAGAGCTGGCTGCATC	156
QY	201	CysThrValSerAspHisIleArgThrHisGluGlnThrThrAlaAlaGluArgGlnThr	220
DB	155	TGCACAGATTTCCGATCATCATCACACAGAGTCCGCTACACCAGAAGAGCGCCAGACA	96
QY	221	ThrPheAsnAspMetIleLysIleAlaLeuGluSerValLeu	234
DB	95	TCATTCAGAACATGATCAAGATCGCTCTTGAAGCTACATTTG	54

[illegible]

Tel: 301-530-9319
Fax: 301-838-0208
Email: carlton@tigr.org
Seq primer: lambda Triplex2.

FEATURES

source
1. .769
/organism="Trichomonas vaginalis"
/mol_type="mRNA"
/strain="T1"
/db_xref="taxon:5722"
/clone="TVTA905"
/note="Vector: Lambda Triplex2; Site 1: SfiA; Site 2: SfiB; T. vaginalis strain T1 library constructed from cDNA, made in lambda Triplex2. Inserts cloned unidirectionally in the SfiA and SfiB sites. Mass excision of library produced inserts in pTriplex2 plasmid. Inserts sequenced from both 5' and 3' ends using Triplex2 sequencing primer and polydT 24 bp primer respectively."

ORIGIN

Alignment Scores:
Pred. No.: 3.16e-75 Length: 769
Score: 681.00 Matches: 134
Percent Similarity: 75.21% Conservative: 42
Best Local Similarity: 57.26% Mismatches: 58
Query Match: 55.59% Indels: 0
DB: 7 Gaps: 0

US-10-035-300A-2 (1-239) x CV204676 (1-769)

```
Qy 1 MetAlaThrProHisIleAsnAlaGluMetGlyAspPheAlaAspValValLeuMetPro 20
Db 757 ATGGCAACACCCCACTACTCGCAAGGTGGCGATTTCGCTGAGACAGTTCATGTGC 698
Qy 21 GlyAspProLeuArgAlaLysTyrIleAlaGluThrPheLeuGluAspAlaArgGluVal 40
Db 697 GCGCATCCACTCCGTGCGCAAGCTCATTCGTGATAAATCTTTCAGAACGCTAAGCAAGTC 638
Qy 41 AsnAsnValArgGlyMetLeuGlyPheThrGlyThrTyrLysGlyArgLysIleSerVal 60
Db 637 AACTCTCTTCGTGAATGCTAGGTATTCAGGACATACAGGCAACCAACCACTCTCTGTT 578
Qy 61 MetGlyHisGlyValGlyIleProSerCysSerIleTyrThrLysGluLeuIleThrAsp 80
Db 577 ATGGGTATGGCATGGGTATCCCATCCATTCATTCACGCTGAGAACTCTACAAGTT 518
Qy 81 PheGlyValLysLysIleIleArgValGlySerCysGlyAlaValLeuProHisValLys 100
Db 517 TACAAGGTTAAGACCATCATCCGTGCGCAGTTCGCGCACAGTTGATCCAAATGTTCCAC 458
Qy 101 LeuArgAspValValIleGlyMetGlyAlaCysThrAspSerLysValAsnArgIleArg 120
Db 457 GTCGCGCATGTCGTCATGTGCACAGCATCTGGCACAGATTCAAATGTTAACAAGATCGC 398
Qy 121 PheLysAspHisAspPheAlaIleAlaAspPheAspMetValArgAsnAlaValAsp 140
Db 397 CTTCCTGGCCACGATTTCCAGCTACAGCCAACTTCGAAGTCGTAGTGCCTTAGTTGAA 338
Qy 141 AlaAlaLysAlaLeuGlyIleAspAlaArgValGlyAsnLeuPheSerAlaAspLeuPhe 160
Db 337 TCGGCCAAGGCTCTCAACATCCCAACACAGGTTGGAAGGCGCTACTCAACAGATATCTTC 278
Qy 161 TyrSerProAspGlyGluMetPheAspValMetGluLysTyrGlyIleLeuGlyValGlu 180
Db 277 TACAGCAAGGAACAAGGATTAATAGGCCCTTCTCAGTACCACTTCATCTGCTGCGAA 218
Qy 181 MetGluAlaAlaGlyIleTyrGlyValAlaAlaGluPheGlyAlaLysAlaLeuThrIle 200
Db 217 ATGGATCTGCTGCTCCCTTCCCAATGCTGATTACTATGCGCAAGAGCTGGCTGCATC 158
Qy 201 CysThrValSerAspHisIleArgThrHisGluGlnThrThrAlaAlaGluArgGlnThr 220
```

```
Db 157 TGCACAGTTCGGATCATCATCACACAGAGTCGCTACCGGAAGAGCGCCAGACA 98
Qy 221 ThrPheAsnAspMetIleLysIleAlaLeuGluSerValLeu 234
Db 97 TCATTTCAGAAATGATCAAGATCGCTCTCTTGAAGCTACATTG 56

RESULT 5
CV204675 792 bp mRNA linear EST 16-SEP-2004
LOCUS EST864385 non-normalized T1 cDNA library Trichomonas vaginalis cDNA
DEFINITION clone TVTA905 3' end, mRNA sequence.
ACCESSION CV204675
VERSION 1 GI:52151655
KEYWORDS Trichomonas vaginalis
SOURCE Trichomonas vaginalis
ORGANISM Trichomonas vaginalis
REFERENCE 1 (bases 1 to 792)
AUTHORS Carlton, J.M., Dyall, S., Johnson, P.J. and Fraser, C.M.
TITLE The complete genome sequence of the sexually transmitted parasite Trichomonas vaginalis
JOURNAL Unpublished (2004)
COMMENT Other ESTs: EST864386
Contact: Jane Carlton
Parasite Genomics Group
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-530-9319
Fax: 301-838-0208
Email: carlton@tigr.org
Seq primer: polydT 24 base.
FEATURES
source
1. 792
/organism="Trichomonas vaginalis"
/mol_type="mRNA"
/strain="T1"
/db_xref="taxon:5722"
/clone="TVTA905"
/clone_lib="non-normalized T1 cDNA library"
/note="Vector: Lambda Triplex2; Site 1: SfiA; Site 2: SfiB; T. vaginalis strain T1 library constructed from cDNA, made in lambda Triplex2. Inserts cloned unidirectionally in the SfiA and SfiB sites. Mass excision of library produced inserts in pTriplex2 plasmid. Inserts sequenced from both 5' and 3' ends using Triplex2 sequencing primer and polydT 24 bp primer respectively."

ORIGIN
Alignment Scores:
Pred. No.: 3.29e-75 Length: 792
Score: 681.00 Matches: 134
Percent Similarity: 75.21% Conservative: 42
Best Local Similarity: 57.26% Mismatches: 58
Query Match: 55.59% Indels: 0
DB: 7 Gaps: 0

US-10-035-300A-2 (1-239) x CV204675 (1-792)
Qy 1 MetAlaThrProHisIleAsnAlaGluMetGlyAspPheAlaAspValValLeuMetPro 20
Db 13 ATGGCAACACCCCACTACTCGCAAGGTGGCGATTTCGCTGAGACAGTTCATGTGC 72
Qy 21 GlyAspProLeuArgAlaLysTyrIleAlaGluThrPheLeuGluAspAlaArgGluVal 40
Db 73 GCGCATCCACTCCGTGCGCAAGCTCATTCGTGATAAATCTTTCAGAACGCTAAGCAAGTC 132
Qy 41 AsnAsnValArgGlyMetLeuGlyPheThrGlyThrTyrLysGlyArgLysIleSerVal 60
Db 133 AACTCTCTTCGTGAATGCTAGGTATTCAGGACATACAGGCAACCAACCACTCTCTGTT 192
Qy 61 MetGlyHisGlyValGlyIleProSerCysSerIleTyrThrLysGluLeuIleThrAsp 80
```

Db 193 ATGGGTATGGCATGGGTATCCCATCCATTTCCATCTCTACGCTGAAGAACTCTACAACTT 252

Qy 81 PheGlyValLysLysIleAArgValGlySerCysGlyAlaValLeuProHisVallys 100

Db 253 TACAAGGTTAAGACCATCATCGTGTGGCACTTGGCGCACAGTTGATCCAAATGTTTCA 312

Qy 101 LeuArgAspValValIleGlyMetAlaCysThrAspSerLysValAsnArgIleArg 120

Db 313 GTCCGCGATGTCGTGATTTGTCACAGCATCTGGCACAGATTCAAATGTTTAAACAGATGGC 372

Qy 121 PheLysAspHisAspPheAlaAlaIleAlaAspPheMetValAlaArgAsnAlaValAsp 140

Db 373 CTTCTTGCCACAGATTCCCAAGCTACAGCCCACTTCGAAGTCGTTAGTGCCTTAGTTGAA 432

Qy 141 AlaAlaLysAlaLeuGlyIleAspAlaArgValGlyAsnLeuPheSerAlaAspLeuPhe 160

Db 433 TCCGCCAAGGCTCTCAACATCCCAACACACAGGTTGGAAAGGCTACTCAACAGATATCTTC 492

Qy 161 TyrSerProAspGlyGluMetPheAspValMetGluLysTyrGlyIleLeuGlyValGlu 180

Db 493 TACAGCAAGGAACAGGATTAATAGAGCCCTTGCTCAGTACCACCTTCATTGCTGTGCA 552

Qy 181 MetGluAlaAlaGlyIleTyrGlyValAlaAlaGluPheGlyAlaLysAlaLeuThrIle 200

Db 553 ATGAATCTGCTGCTCCCTTCCCAATCTGATTAAGGCGCAAGAGCTGGTGCATC 612

Qy 201 CysThrValSerAspHisIleAArgThrHisGluGlnThrThrAlaAlaGluArgGlnThr 220

Db 613 TGCACAGTTTCGGATCATCATCACACAGCTCCGCTACAGGCGGCAAGGCGGCGAGACA 672

Qy 221 ThrPheAsnAspMetIleLysIleAlaLeuGluSerValLeu 234

Db 673 TCATTTCCAGAACATGATCAAGATCGCTCTTGAAGCTACATTG 714

RESULT 6

CV205285

LOCUS

DEFINITION

EST864995 non-normalized T1 cDNA library Trichomonas vaginalis cDNA

clone TVTAC60 5' end, mRNA sequence.

CV205285

CV205285.1 GI:52152265

EST.

Trichomonas vaginalis

Trichomonas vaginalis

Trichomonas vaginalis

Trichomonadidae; Trichomonadinae; Trichomonas.

1 (bases 1 to 794)

Carlton,J.M., Dyll,S., Johnson,P.J., and Fraser,C.M.

The complete genome sequence of the sexually transmitted parasite

Trichomonas vaginalis

Unpublished (2004)

Other ESTs: EST864994

Contact: Jane Carlton

Parasite Genomics Group

The Institute for Genomic Research

9712 Medical Center Drive, Rockville, MD 20850, USA

Tel: 301-530-9319

Fax: 301-838-0208

Email: carlton@tigr.org

Seq primer: lambda Triplex2.

Location/Qualifiers

1..794

/organism="Trichomonas vaginalis"

/mol_type="mRNA"

/strain="T1"

/db_xref="taxon:5722"

/clone="TVTAC60"

/note="Vector: Lambda Triplex2; Site 1: SfiI; Site 2: SfiB; T. vaginalis strain T1 library constructed from cDNA, made in lambda Triplex2. Inserts cloned unidirectionally in the SfiI and SfiB sites. Mass excision of library produced inserts in pTriplex2 plasmid. Inserts

FEATURES

source

sequenced from both 5' and 3' ends using Triplex2 sequencing primer and polydT 24 bp primer respectively."

ORIGIN

Alignment Scores: 3.3e-75 Length: 794

Pred. No.: 681.00 Matches: 134

Score: 75.21% Conservative: 42

Percent Similarity: 75.21% Mismatches: 58

Best Local Similarity: 55.26% Indels: 0

Query Match: 55.59%

DB: 7 Gaps: 0

US-10-035-300A-2 (1-239) x CV205285 (1-794)

Qy 1 MetAlaThrProHisIleAsnAlaGluMetGlyAspPheAlaAspValValLeuMetPro 20

Db 13 ATGGCAACACCCCACTCTGCAAGGTTGGCGATTTTCGCTGAGACAGTTCTCATGTGC 72

Qy 21 GlyAspProLeuArgAlaLysTyrIleAlaGluThrPheLeuGluAspAlaArgGluVal 40

Db 73 GCGCATCCACTCCGTCGCCAAGCTCATTTGCTGATACTATCTTGAGAACGCTAAGCAAGTC 132

Qy 41 AsnAsnValArgGlyMetLeuGlyPheThrGlyThrTyrLysGlyArgLysIleSerVal 60

Db 133 AACTCTGTTCTGGAATGCTTTGGTTTACAGGCACATCAAGGCGCAAGGCGCAAGCTCTCTGTT 192

Qy 61 MetGlyHisGlyValGlyIleProSerCysSerIleTyrThrLysGluLeuIleThrAsp 80

Db 193 ATGGGTATGGCATGGGTATCCCATCCATTTCCATCTACGCTGAAGAACTCTACAACGTT 252

Qy 81 PheGlyValLysLysIleAArgValGlySerCysGlyAlaValLeuProHisVallys 100

Db 253 TACAAGGTTAAGACCATCATCGTGTGGCACTTGGCGCACAGTTGATCCAAATGTTTCA 312

Qy 101 LeuArgAspValValIleGlyMetGlyAlaCysThrAspSerLysValAsnArgIleArg 120

Db 313 GTCCGCGATGTCGATTTGTCACAGCATCTGGCACAGATTCAAATGTTTAAACAGAAATGCGC 372

Qy 121 PheLysAspHisAspPheAlaAlaIleAlaAspPheAspMetValArgAsnAlaValAsp 140

Db 373 CTTCTTGCCACAGATTTCACAGCTACAGCCAACTTCGAAGTCGTTAGTGCCTTAGTTGAA 432

Qy 141 AlaAlaLysAlaLeuGlyIleAspAlaArgValGlyAsnLeuPheSerAlaAspLeuPhe 160

Db 433 TCCGCCAAGGCTCTCAACATCCCAACACACAGGTTGGAAAGGCGCTACTCAACAGATATCTTC 492

Qy 161 TyrSerProAspGlyGluMetPheAspValMetGluLysTyrGlyIleLeuGlyValGlu 180

Db 493 TACAGCAAGGAACAGGATTAATAGAGCCCTTGCTCAGTACCACCTTCATTGCTGTGCA 552

Qy 181 MetGluAlaAlaGlyIleTyrGlyValAlaAlaGluPheGlyAlaLysAlaLeuThrIle 200

Db 553 ATGAATCTGCTGCTCCCTTCCCAATCTGATTAAGGCGCAAGAGCTGGTGCATC 612

Qy 201 CysThrValSerAspHisIleAArgThrHisGluGlnThrThrAlaAlaGluArgGlnThr 220

Db 613 TGCACAGTTTCGGATCATCATCACACAGCTCCGCTACAGGCGGCAAGGCGGCGAGACA 672

Qy 221 ThrPheAsnAspMetIleLysIleAlaLeuGluSerValLeu 234

Db 673 TCATTTCCAGAACATGATCAAGATCGCTCTTGAAGCTACATTG 714

RESULT 7

CV205285

LOCUS

DEFINITION

EST864995 non-normalized T1 cDNA library Trichomonas vaginalis cDNA

clone TVTAC60 5' end, mRNA sequence.

CV205285

CV205285.1 GI:50408335

EST.

Trichomonas vaginalis

Trichomonas vaginalis

Trichomonadidae; Trichomonadinae; Trichomonas.

1 (bases 1 to 794)

Carlton,J.M., Dyll,S., Johnson,P.J., and Fraser,C.M.

The complete genome sequence of the sexually transmitted parasite

Trichomonas vaginalis

Unpublished (2004)

Other ESTs: EST864994

Contact: Jane Carlton

Parasite Genomics Group

The Institute for Genomic Research

9712 Medical Center Drive, Rockville, MD 20850, USA

Tel: 301-530-9319

Fax: 301-838-0208

Email: carlton@tigr.org

Seq primer: lambda Triplex2.

Location/Qualifiers

1..794

/organism="Trichomonas vaginalis"

/mol_type="mRNA"

/strain="T1"

/db_xref="taxon:5722"

/clone="TVTAC60"

/note="Vector: Lambda Triplex2; Site 1: SfiI; Site 2: SfiB; T. vaginalis strain T1 library constructed from cDNA, made in lambda Triplex2. Inserts cloned unidirectionally in the SfiI and SfiB sites. Mass excision of library produced inserts in pTriplex2 plasmid. Inserts

CV205285

LOCUS

DEFINITION

EST864995 non-normalized T1 cDNA library Trichomonas vaginalis cDNA

clone TVTAC60 5' end, mRNA sequence.

CV205285

CV205285.1 GI:50408335

EST.

Trichomonas vaginalis

Trichomonas vaginalis

Trichomonadidae; Trichomonadinae; Trichomonas.

1 (bases 1 to 794)

Carlton,J.M., Dyll,S., Johnson,P.J., and Fraser,C.M.

The complete genome sequence of the sexually transmitted parasite

Trichomonas vaginalis

Unpublished (2004)

Other ESTs: EST864994

Contact: Jane Carlton

Parasite Genomics Group

The Institute for Genomic Research

9712 Medical Center Drive, Rockville, MD 20850, USA

Tel: 301-530-9319

Fax: 301-838-0208

Email: carlton@tigr.org

Seq primer: lambda Triplex2.

Location/Qualifiers

1..794

/organism="Trichomonas vaginalis"

/mol_type="mRNA"

/strain="T1"

/db_xref="taxon:5722"

/clone="TVTAC60"

/note="Vector: Lambda Triplex2; Site 1: SfiI; Site 2: SfiB; T. vaginalis strain T1 library constructed from cDNA, made in lambda Triplex2. Inserts cloned unidirectionally in the SfiI and SfiB sites. Mass excision of library produced inserts in pTriplex2 plasmid. Inserts

Trichomonadidae; Trichomonadinae; Trichomonas.

1 (bases 1 to 721)
Zhou, Y., Shu, W.M., Huang, S.C.C., Huang, K.Y. and Tang, P.

Analysis of Gene Expression Profile in Trichomonas vaginalis by EST
Sequencing

Unpublished (2003)

Contact: Tang, P.

Molecular Regulation and Bioinformatics Laboratory, College of
Medicine

Chang Gung University

259 Wenhu 1st. Road, Kweihsan, Taoyuan 333, Taiwan

Tel: +886 3 3283016 EXT5136

Fax: +886 3 3283031

Email: petang@mail.cgu.edu.tw

PCR Primers

FORWARD: T7

BACKWARD: T3

Seq primer: T3.

Location/Qualifiers

1..721

/organism="Trichomonas vaginalis"

/mol_type="mrna"

/db_xref="taxon:5722"

/cell_line="ATCC30236"

/dev_stage="Trophozoites at mid-log phase"

/lab_hosts="XL1 Blue-MRF"

/clone_lib="TV30236 PT cDNA Library"

/notes="Vector: lambda ZAP-Express (Stratagene); Site_1:

EcoRI; Site_2: XhoI"

ORIGIN

Alignment Scores:
Pred. No.: 1..28e-73 Length: 721
Score: 668.00 Matches: 132
Percent Similarity: 75.00% Conservative: 36
Best Local Similarity: 58.93% Indels: 0
Query Match: 54.53% Gaps: 0
DB: 7

US-10-035-300A-2 (1-239) x COS77895 (1-721)

Qy 11 GlyAspPheAlaAspValValLeuMetProGlyAspProLeuArgAlaLysTyrIleAla 30

Db 44 GCGGATTCGTCGAAACAGTCCTCATGTGCGGTGATCCACTCCGCGCTAAGCTCATGTCT 103

Qy 31 GluThrPheLeuGluAspAlaArgGluValAsnValArgGlyMetLeuGlyPheThr 50

Db 104 GAGACATATCTGAAATCCAAAGCTTGTCAACAATGTTCTGGCATTTCAAGGCTACACC 163

Qy 51 GlyThrThrLysGlyArgGlyIleSerValMetGlyHisGlyValGlyIleProSerCys 70

Db 164 GGCACATACAAAGGAAGCAACATCTGTGATGGCCCATGTATGGCTTGGCCATCATC 223

Qy 71 SerIleThrLysGluLeuIleThrAspPheGlyValLysLysIleIleArgValGly 90

Db 224 TGCATCTATCGAGGAGCTTTACTCCACATACAAAGCTCAAGCAATCATCGTGTGGT 283

Qy 91 SerCysGlyAlaValLeuProHisValLysLeuArgAspValValIleGlyMetGlyAla 110

Db 284 ACATGCGGCGCAATTTGATGGATGGACATCCACACAGTGTATCTTTCACCTCTGCT 343

Qy 111 CysThrAspSerLysValAsnArgIleArgPheLysAspHisAspPheAlaAlaIleAla 130

Db 344 GGTACAACTCCAGATCAACAGATCCGCTTATGGATCAGCATATTCAGCCACAGCA 403

Qy 131 AspPheAspMetValArgAsnAlaValAspAlaAlaLysAlaLeuGlyIleAspAlaArg 150

Db 404 TCTTTCGATGTGTTTTCGCGCTTAGTTGATGCTGCTGAAGCACTCAACATCCAGCTTAAG 463

Qy 151 ValGlyAsnLeuPheSerAlaAspLeuPheThrSerProAspGlyGluMetPheAspVal 170

Db 464 GTCGGTAAGGGATTTCTCAACAGATCTCTTCAAAATCCCAAAACCGAACTCGCACAGCTC 523

Qy 171 MetGluLysTyrGlyIleLeuGlyValGluMetGluAlaAlaGlyIleTyrGlyValAla 190

Db 524 ATGAACAGTTCACATCTCTCGCTGTGTAATGAATCTGCTGCTCTTCCCAATTGCT 583

Qy 191 AlaGluPheGlyAlaLysAlaLeuThrIleCysThrValSerAspHisIleArgThrHis 210

Db 584 GACCTTTATGGCGCAAGAGCTGCTGCATCTGCACAGTTTCAGATCATCATCTCCACCAT 643

Qy 211 GluGlnThrThrAlaAlaGluArgGlnThrPheAsnAspMetLleLysIleAlaLeu 230

Db 644 GAAAAAACACACGCCGAAGAACGCCAGAACTCTTCCAAAACATGACGAAAAATCGCACTT 703

Qy 231 GluSerValLeu 234

Db 704 GAAGCAGCAATC 715

RESULT 8

BH770690/c

LOCUS

DEFINITION

subsp. cremoris genomic, genomic survey sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

1 (bases 1 to 1589)

Lactococcus lactis subsp. cremoris

Lactococcus lactis subsp. cremoris

Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;

Lactococcus.

1 (bases 1 to 1589)

Bolotin, A., Ehrlich, S.D. and Sorokin, A.

Studies of genomes of dairy bacteria Lactococcus lactis

Sci. Aliments (2002) In press

Contact: Sorokin, A

Genetique Microbienne

INRA

CRJ INRA, Domaine de Vilvert, 78352 Jouy en Josas cedex, France

Tel: 33 1 34 65 25 16

Fax: 33 1 34 65 25 21

Email: sorokine@jouy.inra.fr

best homologue in strain IL1403 is deoD (99%)

Class: shotgun

High quality sequence start: 30

High quality sequence stop: 1561.

Location/Qualifiers

1..1589

/organism="Lactococcus lactis subsp. cremoris"

/mol_type="genomic DNA"

/strain="MG1363"

/sub_species="cremoris"

/db_xref="taxon:1359"

/clone_lib="MG1363 Random Sequence Tag Library"

/notes="Vector: pSGW02; Site_1: SmaI; Library of

chromosomal fragments of L.lactis strain MG1363 was

prepared by partial AluI digestion or by sonication."

ORIGIN

Alignment Scores:

Pred. No.: 2.1e-63 Length: 1589

Score: 591.00 Matches: 120

Percent Similarity: 70.75% Conservative: 30

Best Local Similarity: 56.60% Mismatches: 58

Query Match: 48.24% Indels: 4

DB: 8 Gaps: 2

US-10-035-300A-2 (1-239) x BH770690 (1-1589)

Qy 1 MetAlaThrProHisIleAsnAlaGluMetGlyAspPheAlaAspValValLeuMetPro 20

Db 646 ATCCCAACACCATATCGAAGCTCAAAAGCGGAAATCCAGATAAAATCTCTTACCA 587

Qy 21 GlyAspProLeuArgAlaLysTyrIleAlaGluThrPheLeuGluAspAlaArgGluVal 40

Db 586 GGAGATCCACTTCGCGCAAAATTTATCGCAGAAAACCTTCTTGAAGATGCTGTGCAATT 527

Qy	41	Asn	Asn	Val	Arg	Gly	Me	Leu	Gly	Phe	Thr	Gly	Thr	Tyr	Lys	Gly	Arg	Lys	Ile	Ser	Val	60
Db	526	AAC	CAAG	TTCC	GGAA	TCTT	GGCT	TTT	AT	CGAA	CTT	AC	AGG	GC	AT	CT	CA	AGG	GC	AT	CGT	467
Qy	61	Met	Gly	His	Gly	Val	Gly	Ile	Pro	Ser	Cys	Ser	Ile	Tyr	Thr	Lys	Glu	Leu	Ile	Thr	Asp	80
Db	466	ATG	GA	AC	AG	AA	TGG	GA	AT	TCC	TT	CA	AT	CT	CA	AT	CT	CA	AT	CT	CA	407
Qy	81	Phe	Gly	Val	Lys	Ile	Ile	Arg	Val	Gly	Ser	Cys	Gly	Ala	Val	Leu	Pro	His	Val	Lys	100	
Db	406	TAT	GG	CG	TT	AA	AG	AT	TA	AT	TTC	GT	GG	AT	CT	GT	GC	TA	AT	CT	GC	347
Qy	101	Leu	Arg	Asp	Val	Val	Ile	Gly	Met	Gly	Ala	Cys	Thr	Asp	Ser	Lys	Val	Asn	Arg	Ile	Arg	120
Db	346	ATT	CG	CA	CA	CT	TG	T	CG	T	GG	AT	CT	CA	CA	CA	CA	CA	CA	CA	CA	287
Qy	121	Phe	Lys	Asp	His	Asp	Phe	Ala	Ala	Ile	Ala	Asp	Phe	Asp	Met	Val	Arg	Asn	Ala	Val	Asp	140
Db	286	TTT	CT	GA	TTT	GA	TTT	CCC	ACA	AA	TTC	TA	TTT	GA	TTT	GA	TTT	GA	TTT	GA	TTT	227
Qy	141	Ala	Ala	Lys	Ala	Leu	Gly	Ile	Asp	Ala	Arg	Val	Gly	Asn	Leu	Phe	Ser	Ala	Asp	Leu	Phe	160
Db	226	ATT	G	CA	AA	AG	ACC	TT	TG	GA	AT	TACA	CA	CT	CAC	GT	TG	GA	AT	AT	TCT	167
Qy	161	Tyr	---	Ser	Pro	Asp	Gly	Glu	Met	Phe	Asp	Val	Met	Glu	Lys	Tyr	Gly	Ile	Leu	Gly	Val	179
Db	166	TAT	GG	TG	GC	AC	GAT	GC	AG	TT	TAAA	-----	GT	TG	AA	AA	CT	CG	GT	GT	GA	116

LOCUS	878 bp	DNA	linear	GSS 03-NOV-2000
ENTB4777R	Entamoeba histolytica	Sheared	DNA	Entamoeba histolytica
DEFINITION	genomic, genomic survey sequence.			

ACCESSION	AZ533877
VERSION	AZ533877.1
KEYWORDS	GI:11089945
SOURCE	GSS.
ORGANISM	Entamoeba histolytica
REFERENCE	Entamoeba histolytica
AUTHORS	Eukaryota; Entamoebidae; Entamoeba.
TITLE	1 (bases 1 to 878)
JOURNAL	Loftus B., Van Aken S. and Fraser, C.
COMMENT	Determination of clone end sequences from Entamoeba histolytica HMI:MSS sheared DNA library Unpublished (2000) Contact: Brendan J Loftus

Department of Eukaryotic Genomics
The Institute for Genomic Research
9112 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Fax: 301 838 3543
Email: bjloftus@tigr.org
Clones are derived from the Entamoeba histolytica HML1.MSS sheared
DNA library M13-Reverse
Seq primer: M13-Reverse
Class: Shotgun
High quality sequence start: 14
High quality sequence stop: 847.

```

FEATURES
source
Location/Qualifiers
1..878
/organism="Entamoeba histolytica"
/mol_type="genomic DNA"
/strain="HM1:IMSS"
/db_xref="taxon:5759"
/collection="Entamoeba histolytica Sheared DNA"

```

/note=Vector; Site_1: Bst I; Constructed at The Institute for Genomic Research (TIGR), Rockville, MD. Genomic DNA isolated from broth cultures of *E. histolytica* using a method described by Clark and Diamond (Clark, C.G., and Diamond, L.S. (1993) *Entamoeba histolytica*: a method for isolate identification. Exp. Parasitol. 71:450.). The DNA was mechanically sheared to give a tight size distribution (~2 kb). The λ i method used for the library construction is described in detail in Smith, H.O. and Venter, J.C. (making small insert libraries for whole genome shotgun sequencing projects. In Genome Sequencing: A Practical Approach, eds. M. Vaudin and B. Barrell, Oxford University Press, 1999)."

ORIGIN

Alignment Scores:	
Pred. No.:	3.32e-44
Score:	436.50
Percent Similarity:	63.68%
Best Local Similarity:	46.32%
Query Match:	35.63%
DB:	8
Length:	878
Matches:	88
Conservative:	33
Mismatches:	68
Indels:	
Gaps:	1

US-10-035-300A-2 (1-239) x AZ533877 (1-878)

Qy	41	AsnAsnValArgGlyMetLeuGlyPheThrGlyThrTyrTyrLysGlyArgLysIleSerVal	60
Db	878	AATGGAGTAAAGAGGGGCAATTAGGATATCTACTGGATATTCACAAAGAGAAAGAGGTTTCTATC	819

Qy 61 MetGlyHisGlyValGlyIleProSerCysSerIleTyrThrLysGluLeuIleThrAsp 80

DB 818 AAAGCTCATGGTATGGGTATACCATCTATTGGAAATTTATGCATATGAATTATTTAATTC 759

QY 81 PheGlyValLysIleIleArgValGlySerCysGlyAlaValLeuProHisValLys 100

DB 758 TATGGAGTAAAGAGAAATATATAGAAATGGTTCCTGCTGGAGCTAATAGCTCTAATAATAGAT 699

[illegible]

121 pheI.vsaAsnHisAsnGpeA]aA]aIleA]aaAsnGpeAsnMetValArgAsnA]aValAsn 140

Db
638 AAATCCCAGGAACATTTCTGCCCCAATAGCTGATTATGAGTTAGTTCATAAAGCTGTTGAA 579

Qy 141 AlaAlaLysAlaLeuGlyIleAspAlaArgValGlyAsnLeuPheSerAlaAspLeuphe 160

Db 578 GCTGTTGAAAGATTGGAGCTAAGTATGCAGTAGGGAATGTTTTTCTAGTGATTGTTTC 519

Qy
161 TyrSerProAspGlyGluMetPheAspValMetGluLysTyrGlyIleLeuGlyValGlu 180
Qy
161 TyrSerProAspGlyGluMetPheAspValMetGluLysTyrGlyIleLeuGlyValGlu 180

DB 518 TATGAT---GATTCAGTTCATTAAAAGAAATGGGATAAGATGGGAGTACATTGCTGTAGAA 462

[illegible]

1000

401 TGTCTATTCTGATTGTCCATTTAAGGAAACTGCTCTCTCAGCTGAGAGAACAAACC 342

Qy 221 ThrPheAsnAspMetIleLysIleAlaLeu 230

Db 341 AAATTTACACAAATGATGGAAAGTTGCACTG 312

RESULT 10
PH138074

LOCUS	BH138974	921 bp	DNA	linear	GSS 07-AUG-200
DEFINITION	ENTOV94TF	Entamoeba histolytica	Sheared DNA	Entamoeba histolytica	

genomic, genomic survey sequence.

KEYWORDS
GSS.

SOURCE Entamoeba histolytica
 ORGANISM Entamoeba histolytica
 Eukaryota; Entamoebidae; Entamoeba.
 1 (bases 1 to 921)
 REFERENCE Loftus,B., Wang,Z., Van Aken,S. and Fraser,C.
 AUTHORS Determination of clone end sequences from Entamoeba histolytica
 TITLE HMI:IMSS sheared DNA library (2001)
 JOURNAL Unpublished (2001)
 COMMENT Contact: Brendan J Loftus
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0208
 Fax: 301 838 3543
 Email: bjloftus@tigr.org
 Clones are derived from the Entamoeba histolytica HMI:IMSS sheared
 DNA library
 Seq primer: M13-Forward
 Class: shotgun
 High quality sequence start: 2
 High quality sequence stop: 563.
 FEATURES
 source
 1..921
 /organism="Entamoeba histolytica"
 /mol_type="genomic DNA"
 /strain="HMI:IMSS"
 /db_xref="taxon:5759"
 /clone_lib="Entamoeba histolytica Sheared DNA"
 /notes="Vector: pHSI; Site 1: Bst I; Constructed at The
 Institute for Genomic Research (TIGR), Rockville, MD.
 Genomic DNA isolated from broth cultures of E. histolytica
 using a method described by Clark and Diamond (Clark,
 C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a
 method for isolate identification. Exp. Parasitol.
 77:450.). The DNA was mechanically sheared to give a
 tight size distribution (~2 kb). The v + i method used for
 the library construction is described in detail in Smith,
 H.O. and Venter, J.C. (Making small insert libraries for
 whole genome shotgun sequencing projects. In Genome
 Sequencing: A Practical Approach, eds. M. Vaudin and B.
 Borell, Oxford University Press, 1999)."

ORIGIN

Alignment Scores:
 Pred. No.: 3,99e-44 Length: 921
 Score: 436.00 Matches: 108
 Percent Similarity: 65.18% Conservative: 38
 Best Local Similarity: 48.21% Mismatches: 74
 Query Match: 35.59% Indels: 7
 DB: 8 Gaps: 3

US-10-035-300A-2 (1-239) x BH138974 (1-921)

QY 3 ThrProHisIleAsnAlaGluMetGlyAspPheAlaAspValValLeuMetProGlyAsp 22
 Db 130 ACTCCACATAATGGTGTAAATACGGAGAGATTGCTGAAACTGCTCTTATGGCTGGTAT 189
 QY 23 ProLeuArgAlaLysTyrlleAlaGluThrPheLeuGluAspAlaArgGluValAsnAsn 42
 Db 190 CCATTAAGAGATAAACTCTTTCGACAGACACTATTTCGACTGATGTGTTCAATAATAAGT 249
 QY 43 ValArgGlyMetLeuGlyPheThrGlyThrTyrlsGlyArgLysIleSerValMetGly 62
 Db 250 GTTAGAGGACGATGAGATATATCTGGATATATACAGGAGTGAAGTATCAGTTCAAGCT 309
 QY 63 HisGlyValGlyIleProSerCysSerIleTyrlsGlyLeuLeuIleThrAspPheGly 82
 Db 310 CATGGTATGGTATGCCATCTATTGGAAATTTATGCATATCAATATTATTAATTTCTATGGA 369
 QY 83 ValLysLysIlelleArgValGlySerCysGlyAlaValleuProHisValLysLeuArg 102
 Db 370 GTAAGAGAAATTATTAGAAATGGTTCGTCGGAGCTTTTCATGAAAGTTTAAAACTTGA 429

QY 103 AspValValIleGlyMetGlyAlaCysThrAspSerLysValAsnArgIleAArgPheLys 122
 Db 430 GATATTGTTTGAATGGAGCATGTTATGATTCCTCAATTTGAAAGACAAATATGATATT 489
 QY 123 AspHisAspPheAlaAlaIleAlaAspPheAspMetValArgAsnAlaValAspAlaAla 142
 Db 490 CCAGGTAAATACTCATGCTATTGCTGATTTCCAACTTTGCAGAGAAGCAGTTGATGCAGCT 549
 QY 143 LysAlaLeuGlyIleAspAlaArgValGlyAsnLeuPheSerAlaAspLeuPheTyrSer 162
 Db 550 GAAAACTCGATATATATATATTTATTTCTCTAATATTATTTCTTATGAT 608
 QY 163 ProAspGlyGluMetPheAspValMetGluLysTyrlsGlyIleLeuGlyValGluMetGlu 182
 Db 609 --GATGGAGATCACTCCGGAGCATCGAAAAGATGGAGTACTTGTCTAGAAAT-GAA 664
 QY 183 AlaAlaGlyIleTyrlsGlyValAlaAlaGluPheGlyAlaLys-AlaLeuThrIleCysTh 202
 Db 665 GCAGCTGCTCTTTATATATGATTGCAGCAGACAC--GTAACACAGCAGCTTTGTATGGTAAC 721
 QY 202 rValSerAsp---HisIleArgThrHisGluGlnThrThrAlaAlaGluArgGlnThrTh 221
 Db 722 TATTTCTGATCTTTGTTGTAAGAGTGGGAAAAATG-ACAGCCGAAAGAAAGAACTAAT 780
 QY 221 rPheAsnAsp 224
 Db 781 TTTCTCCAAT 790

RESULT 11
 AZ680406/c

LOCUS ENTGM407F Entamoeba histolytica Sheared DNA linear GSS 14-DEC-2000
 DEFINITION genomic, genomic survey sequence.

ACCESSION AZ680406
 VERSION AZ680406.1 GI:11817552
 KEYWORDS GSS.

SOURCE

Entamoeba histolytica

Eukaryota; Entamoebidae; Entamoeba.

1 (bases 1 to 863)

Loftus,B., Van Aken,S. and Fraser,C.

Determination of clone end sequences from Entamoeba histolytica

HMI:IMSS sheared DNA library

REFERENCE

Unpublished (2000)

Contact: Brendan J Loftus

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0208

Fax: 301 838 3543

Email: bjloftus@tigr.org

Clones are derived from the Entamoeba histolytica HMI:IMSS sheared

DNA library

Seq primer: M13-Forward

Class: shotgun

High quality sequence start: 26

High quality sequence stop: 769.

Location/Qualifiers

FEATURES

source

1..863
 /organism="Entamoeba histolytica"
 /mol_type="genomic DNA"
 /strain="HMI:IMSS"
 /db_xref="taxon:5759"
 /clone_lib="Entamoeba histolytica Sheared DNA"
 /notes="Vector: pHSI; Site 1: Bst I; Constructed at The
 Institute for Genomic Research (TIGR), Rockville, MD.
 Genomic DNA isolated from broth cultures of E. histolytica
 using a method described by Clark and Diamond (Clark,
 C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a
 method for isolate identification. Exp. Parasitol.
 77:450.). The DNA was mechanically sheared to give a
 tight size distribution (~2 kb). The v + i method used for
 the library construction is described in detail in Smith,

H.O. and Venter, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. In *Genome Sequencing: A Practical Approach*, eds. M. Vaudin and B. Barrell, Oxford University Press, 1999). "

ORIGIN

Alignment Scores:

Pred. No.:	2.42e-43	Length:	8
Score:	429.50	Matches:	3
Percent Similarity:	65.14%	Conservative:	31
Best Local Similarity:	47.43%	Mismatches:	60
Query Match:	35.06%	Indels:	1
DB:	8	Gaps:	1

US-10-035-300A-2 (1-239) x AZ680406 (1-863)

Qy	3	ThrProHisIleAsnAlaGluMetClyAspPheAlaAspValValLeuMetProGlyAsp	22
Db	545	ACTCCACACAAACACGCCTAAATAGCTGATTTTCTGATACAGTGTGTAGCTGCTGAT	486
Qy	23	ProLeuArgAlaIleValTyrIleAlaGluThrPheLeuGluAspAlaArgGluValaAsn	42
Db	485	CCACTAAGATGTAATAATTTGTTGCACACTATTATTGATCTGATGTTGTTCAAGTCATGGA	426
Qy	43	ValArgGlyMetLeuGlyPheThrGlyThrTyrLysGlyArgLysIleSerValMetGly	62
Db	425	GTAAGAGGGGCATTAGGATATCTGGATATTACAAAGAAAGGTTCTTCTTCAAGCT	366
Qy	63	HisGlyValGlyIleProSerCysSerIleTyrThrLysGluLeuIleThrAspPheGly	82
Db	365	CATGGTATGGGTATACCATCTATTGGCAATTATGTCATATGAAATTATTAATTTCTATGGA	306
Qy	83	ValLysLysIleIleArgValGlySerCysGlyAlaValLeuProHisValLysLysLeuArg	102
Db	305	GTAAGAGAAATTATTAGAATTGGTTCTGCTGGAGCTATTAGTCTAAATTAGATATTGGA	246
Qy	103	AspValValIleGlyMetClyAlaCysThrAspSerLysValAsnArgIleArgPheLys	122
Db	245	GATATTGTTCATGACAACTGGTGGCTGTGTCATATAGTAATTTTATGCAACAATATAA	186
Qy	123	AspHisAspPheAlaIleAlaAspPheAspMetValArgAsnAlaValAspAlaAla	142
Db	185	CCAGAACAACTTGCCTCAATAGCTGATTATGAGTTAGTTTCATTAAGCTGTTGAAGCTGTT	126
Qy	143	LysAlaLeuGlyIleAspAlaArgValGlyAsnLeuPheSerAlaAspLeuPheTyrSer	162
Db	125	GAAAGATTGGAGCTAAGATGATGCATAGGGAATGTTTTTCTAGTGATGTTCTTCTATGAT	66
Qy	163	ProAspGlyGluMetPheAspValMetCgluLysTyrGlyIleLeu	177
Db	65	---GATTCAAGTTCAATTAAGAATGGGATGAAGATGGGAGTACTT	24

RESULT 12

AB050212.1	BH166036	861 bp	DNA	linear	GSS 24-SEP-2001
LOCUS	ENT57607F	Entamoeba histolytica	Sheared DNA	Entamoeba histolytica	
DEFINITION	genomic, genomic survey sequence.				
ACCESSION	BH166036				
VERSION	BH166036.1	GI:15739474			
KEYWORDS	GSS.				
SOURCE	Entamoeba histolytica				
ORGANISM	Entamoeba histolytica				
REFERENCE	Eukaryota; Entamoebidae; Entamoeba.				
AUTHORS	1 (bases 1 to 861)				
TITLE	Loftus, B., Wang, Z., Van Aken, S. and Fraser, C.				
	Determination of clone end sequences from Entamoeba histolytica				
	RM1:IMSS sheared DNA library (2001)				
JOURNAL	Unpublished (2001)				
COMMENT	Contact: Brendan J Loftus				
	Department of Eukaryotic Genomics				
	The Institute for Genomic Research				
	9712 Medical Center Dr., Rockville, MD 20850, USA				
	Tel: 301 838 0208				

```

AZ675191
LOCUS      AZ675191      864 bp      DNA      linear      GSS 14-DEC-2000
DEFINITION ENT59TR Entamoeba histolytica Sheared DNA Entamoeba histolytica
            genomic, genomic survey sequence.
ACCESSION  AZ675191
VERSION    AZ675191.1  GI:11812337
KEYWORDS   GSS.
SOURCE     Entamoeba histolytica
ORGANISM   Entamoeba histolytica
            Eukaryota; Entamoebidae; Entamoeba.
REFERENCE  1 (bases 1 to 864)
AUTHORS    Loftus,B., Van Aken,S. and Fraser,C.
TITLE      Determination of clone end sequences from Entamoeba histolytica
            HMI:IMSS sheared DNA library
JOURNAL    Unpublished (2000)
COMMENT    Contact: Brendan J Loftus
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0208
            Fax: 301 838 3543
            Email: bjloftus@tigr.org
            Clones are derived from the Entamoeba histolytica HMI:IMSS sheared
            DNA library
            Seq primer: M13-Reverse
            Class: shotgun
            High quality sequence start: 21
            High quality sequence stop: 808.

FEATURES             source
    source
        1..864
            /organism="Entamoeba histolytica"
            /mol_type="genomic DNA"
            /strain="HMI:IMSS"
            /db_xref="taxon:5759"
            /clone_lib="Entamoeba histolytica Sheared DNA"
            /note="Vector: pHOS1; Site 1: Bst I; Constructed at The
            Institute for Genomic Research (TIGR), Rockville, MD.
            Genomic DNA isolated from broth cultures of E. histolytica
            using a method described by Clark and Diamond (Clark,
            C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a
            method for isolate identification. Exp. Parasitol.
            77:450.). The DNA was mechanically sheared to give a
            tight size distribution (~2 kb). The v + i method used for
            the library construction is described in detail in Smith,
            H.O. and Venter, J.C. (Making small insert libraries for
            whole genome shotgun sequencing projects. In Genome
            Sequencing: A Practical Approach, eds. M. Vaudin and B.
            Barrell, Oxford University Press, 1999)."
```

```

    Alignment Scores:      4.69e-35      Length:      864
    Pred. No.:            364.00      Matches:      79
    Score:                 67.32%      Conservative:  24
    Best Local Similarity: 51.63%      Mismatches:   49
    Query Match:          29.71%      Indels:      2
    DB:                     8          Gaps:         0

US-10-035-300A-2 (1-239) x AZ675191 (1-864)

Qy      3 ThrProHisIleAsnAlaGluMetGlyAspPheAlaAspValValLeuMetProGlyAsp 22
Db      409 ACTCCACATATGCTGCTAATAACGAGAGATTGCTGAACTGTCCTATGGCTGGTAT 468
Qy      23 ProLeuArgAlaLysTyrlleAlaGluThrPheLeuGluAspAlaArgGluValAsn 42
Db      469 CCATTAAAGAGPAAAACTCTTCGACACACTTATTTCGATGTTGTTCAATAATAGT 528
Qy      43 ValArgGlyMetLeuGlyPheThrGlyThrTyrlsGlyArgGlyIleSerValMetGly 62
Db      529 GTTAGAGAGCAGGAGGATGATCTGGATATTACAGGGGATGAGTATCATGTTCAAGCT 588
Qy      63 HisGlyValGlyIleProSerCysSerIleTyrlsGluLeuIleThrAspPheGly 82

ORIGIN
Alignment Scores:      1.84e-32      Length:      865
Pred. No.:            343.50      Matches:      64
Score:

```

```

Db      589 CATGGTATGGTATGCCATCTATTGGAATTTATGTCATATGAATTTAATTTCTATGGA 648
Qy      83 ValLysLysIleIleArg-ValGlySerCysGlyAlaValLeuProHisValLysLeuAr 102
Db      649 GTAAAGAGAATTATTAGAAATTCGTTCTGCTGAGCCTTTGATGAAAGTTTAAACACTGG 708
Qy      102 GAspValValIleGlyMetGlyAlaCysThrAspSerLysValAsnArgIleArgPheLy 122
Db      709 AGATATTGTTATTGGAATGGAGCATGTTATGATTTCCCAATTTGAAAGACAATATGATAT 768
Qy      122 sAspHisAspPheAlaIleAlaAspPheAspMetValArgAsnAlaValAspAlaAl 142
Db      769 TCCAGGTAAATACTCATGTATTGCTGATTTCCAACTTTGCAGAGAAACAGTTGATGCAGC 828
Qy      142 aLysAlaLeuGlyIleAspAlaArgValGlyAsnLeu 154
Db      829 TGAATAA-ACCTGGTATAGATATAAGGTTAGGAATTTA 864

RESULT 14
LOCUS    AZ669801/c      865 bp      DNA      linear      GSS 14-DEC-2000
DEFINITION ENTLL39TF Entamoeba histolytica Sheared DNA Entamoeba histolytica
            genomic, genomic survey sequence.
ACCESSION  AZ669801
VERSION    AZ669801.1  GI:11806947
KEYWORDS   GSS.
SOURCE     Entamoeba histolytica
ORGANISM   Entamoeba histolytica
            Eukaryota; Entamoebidae; Entamoeba.
REFERENCE  1 (bases 1 to 865)
AUTHORS    Loftus,B., Van Aken,S. and Fraser,C.
TITLE      Determination of clone end sequences from Entamoeba histolytica
            HMI:IMSS sheared DNA library
JOURNAL    Unpublished (2000)
COMMENT    Contact: Brendan J Loftus
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0208
            Fax: 301 838 3543
            Email: bjloftus@tigr.org
            Clones are derived from the Entamoeba histolytica HMI:IMSS sheared
            DNA library
            Seq primer: M13-Forward
            Class: shotgun
            High quality sequence start: 18
            High quality sequence stop: 707.

FEATURES             source
    source
        1..865
            /organism="Entamoeba histolytica"
            /mol_type="genomic DNA"
            /db_xref="taxon:5759"
            /strain="HMI:IMSS"
            /clone_lib="Entamoeba histolytica Sheared DNA"
            /note="Vector: pHOS1; Site 1: Bst I; Constructed at The
            Institute for Genomic Research (TIGR), Rockville, MD.
            Genomic DNA isolated from broth cultures of E. histolytica
            using a method described by Clark and Diamond (Clark,
            C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a
            method for isolate identification. Exp. Parasitol.
            77:450.). The DNA was mechanically sheared to give a
            tight size distribution (~2 kb). The v + i method used for
            the library construction is described in detail in Smith,
            H.O. and Venter, J.C. (Making small insert libraries for
            whole genome shotgun sequencing projects. In Genome
            Sequencing: A Practical Approach, eds. M. Vaudin and B.
            Barrell, Oxford University Press, 1999)."
```

Percent Similarity: 68.95% Conservative: 20
 Best Local Similarity: 52.46% Mismatches: 33
 Query Match: 28.04% Indels: 5
 DB: 8 Gaps: 1

US-10-035-300A-2 (1-239) x AZ669801 (1-865)

Qy 3 ThrProHisIleAenAlaGluMetGlyAspPheAlaAspValValLeuMetProGlyAsp 22
 Db 365 ACTCCACACAAACGCTAAATAGCTGATTGCTGATACAGTGTGTAGTACGCTGAT 306
 Qy 23 ProLeuArgAlaLysTyrIleAlaGluThrPheLeuGluAspAlaArgGluValAsn 42
 Db 305 CCACTAAGATCAATATTGTTGCACACACTTATTGCTGATGTTGTTCAAGTCAATGGA 246
 Qy 43 ValArgGlyMetLeuGlyPheThrGlyThrTyrLysGlyArgLysIleSerValMetGly 62
 Db 245 GTAAGAGGGGATGATGATATGCTGATATCAAGGAAAAAGTTTCTATTCAAGCT 186
 Qy 63 HisGlyValGlyIleProSerCysSerIleTyrThrLysGluLeuIleThrAspPheGly 82
 Db 185 CATGGTATGGGTATACCATCTATTGGAATTTATGATGATGATGATTTAATTTCTATGGA 126
 Qy 83 ValLysLysIleIleArgValGlySerCysGlyAlaValLeuProHisValLysLeuArg 102
 Db 125 GTAAGAGAGAAATTTAGAAATTTGCTGCTGAGCTATTAGCTCTAAATTAGATTTGGA 66
 Qy 103 AspValValIleGlyMetGlyAlaCysThrAspSerLysValAsnArgIleArgPheLys 122
 Db 65 GATATTGTCATAGCACTGGTGTGTCATATAGT-----AATTTATG 21
 Qy 123 AspHis 124
 Db 20 CAACAT 15

RESULT 15
 AZ681446/c
 LOCUS
 DEFINITION ENTMK91TR Entamoeba histolytica Sheared DNA Entamoeba histolytica
 genomic, genomic survey sequence.
 ACCESSION AZ681446
 VERSION AZ681446.1 GI:11818592
 KEYWORDS GSS.

ORGANISM
 Entamoeba histolytica
 Entamoeba histolytica
 Eukaryota; Entamoebidae; Entamoeba.
 1 (bases 1 to 882)
 Loftus, B., Van Aken, S. and Fraser, C.
 Determination of clone end sequences from Entamoeba histolytica
 HMI:IMSS sheared DNA library
 Unpublished (2000)

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 COMMENT
 Contact: Brendan J Loftus
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0208
 Fax: 301 838 3543
 Email: bjoftus@tigr.org
 Clones are derived from the Entamoeba histolytica HMI:IMSS sheared
 DNA library
 Seq primer: M13-Reverse
 Class: shotgun
 High quality sequence start: 14
 High quality sequence stop: 804.

FEATURES
 source
 1..882
 /organism="Entamoeba histolytica"
 /mol_type="genomic DNA"
 /strain="HMI:IMSS"
 /db_xref="taxon:5759"
 /clone_lib="Entamoeba histolytica Sheared DNA"
 /note="Vector: pHOSt; Site 1: Bst I; Constructed at The
 Institute for Genomic Research (TIGR), Rockville, MD.

ORIGIN

Alignment Scores:
 Pred. No.: 8,82e-29 Length: 882
 Score: 314.50 Matches: 68
 Percent Similarity: 60.38% Conservative: 28
 Best Local Similarity: 42.77% Mismatches: 62
 Query Match: 25.67% Indels: 1
 DB: 8 Gaps: 1
 US-10-035-300A-2 (1-239) x AZ681446 (1-882)
 Qy 72 IleTyrThrLysGluLeuIleThrAspPheGlyValLysLysIleIleArgValGlySer 91
 Db 876 ATTTATGATATGAATTTATTTCTATGAGTAAAGAGAAATTTAGATTTGGTTCT 817
 Qy 92 CysGlyAlaValLeuProHisValLysLeuArgAspValValIleGlyMetGlyValaCys 111
 Db 816 GCTGAAGCTATTAGCTCTAAATTAGATATTGAGATATTGTCATAGCAACTGGTCTGT 757
 Qy 112 ThrAspSerLysValAsnArgIleArgPheLysAspHisAspPheAlaIleAlaAsp 131
 Db 756 CATATATGATATTTTATGCAACAATATAAACTCCAGGAACATTTGCCCAATAGCTGAT 697
 Qy 132 PheAspMetValArgAsnAlaValAspAlaLysAlaLeuGlyIleAspAlaArgVal 151
 Db 696 TATGAGTTAGTTTCAATAAGCTGTTTGAAGCTGTTGAAAAGATTGGAGCTTAAGTATGCACTA 637
 Qy 152 GlyAsnLeuPheSerAlaAspLeuPheTyrSerProAspGlyGluMetPheAspValMet 171
 Db 636 GGGAAATGTTTTTCTAGTATGTTTCTATGAT---GATTCAGTTTCATTAAGAAGATGG 580
 Qy 172 GluLysTyrGlyIleLeuGlyValGluMetGluAlaAlaGlyIleTyrGlyValAlaAla 191
 Db 579 GATAAGATGGGAGTACTTGTCTGTAGAAATGAAGCAGCTGTTTATATATGCTATTGCGCA 520
 Qy 192 GluPheGlyAlaLysAlaLeuThrIleCysThrValSerAspHisIleArgThrHisGlu 211
 Db 519 AGACGCGCAACCAAGCTTTGACTATTGTTCTATTTCTGATTCTCCATTTAAGGAAACT 460
 Qy 212 GlnThrThrAlaAlaGluArgGlnThrThrPheAsnAspMetIleLysIleAlaLeu 230
 Db 459 GCTCTCTACGCTGAGAGAAAGAACCAACCAATTTACACAAATGATGGAAGTTGCATG 403

Search completed: July 23, 2005, 04:25:30
 Job time : 3882 secs

This Page Blank (Inten)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 18, 2005, 21:22:05 ; Search time 174 Seconds
(without alignments)
703.373 Million cell updates/sec

Title: US-10-035-300A-2
Perfect score: 1225
Sequence: 1 MATPHINAEMGDFADVLMF.....TTFNDMIKIALESVLLGDKE 239

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : UniProt_03.4

1: uniprot_sprot.*

2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1221	99.7	239	Q7UA16	Q7ua16 shigella fl
2	1221	99.7	258	Q83P00	Q83p00 shigella fl
3	1221	99.7	258	Q8FA51	Q8fa51 escherichia
4	1217	99.3	238	1 DEOD_ECOLI	P09743 escherichia
5	1187	96.9	239	2 Q8Z0J2	Q8z0u2 salmonella
6	1182	96.5	239	2 Q8ZJV7	Q8zjv7 salmonella
7	1115	91.0	239	2 Q66EV7	Q66ev7 yersinia ps
8	1106	90.3	239	2 Q8ZIQ2	Q8zjq2 yersinia pe
9	1106	90.3	241	2 Q8CZM9	Q8czn9 yersinia pe
10	1095	89.4	239	2 Q6D989	Q6d989 erwinia car
11	1085	88.7	238	2 Q8KRT5	Q8krt5 xenorhabdus
12	1085	88.6	239	1 DEOD_KLEPN	Q59482 klebsiella
13	1067	87.1	238	2 Q7N930	Q7n930 photorhabdu
14	1039	84.8	238	2 Q9CL56	Q9cle6 pasteurella
15	1021	83.3	238	2 Q65RA4	Q65ra4 mannheimia
16	1018	83.1	239	2 Q87M25	Q87m25 vibrio para
17	1017	83.0	238	1 Q8DE9	Q4417 haemophilus
18	1012	82.6	239	2 Q8DE9	Q8dbes9 vibrio vuln
19	1012	82.6	254	2 Q7MI41	Q7mi41 vibrio vuln
20	996	81.3	241	2 Q6LUH1	Q6luh1 photobacter
21	994	81.1	240	1 DEOD_ACTPL	P94164 actinobacil
22	993	81.1	241	2 Q9KPM0	Q9kpm0 vibrio chol
23	993	76.2	238	2 Q7VMS8	Q7vms8 haemophilus
24	861	70.3	236	2 Q8EHK0	Q8ekho shewanella
25	858	70.0	236	2 Q6LLA7	Q6lla7 photobacter
26	858	70.0	245	2 Q9KNE2	Q9knb2 vibrio chol
27	853	69.6	236	2 Q87G42	Q87g42 vibrio para
28	834	68.1	236	2 Q8D322	Q8d322 vibrio vuln
29	834	68.1	274	2 Q7MFG6	Q7mfg6 vibrio vuln
30	778	63.5	234	1 DEOD_EUCAI	P57606 buchnera ap
31	747	61.0	236	2 Q7NRT2	Q7nrt2 chromobacte

32	729	59.5	236	2 Q8EKK0	Q8ekko shewanella
33	712	58.1	238	1 DEOD_BUCBP	Q89a58 buchnera ap
34	700	57.1	235	2 Q63DR9	Q63dr9 bacillus ce
35	699	57.1	235	2 Q73B32	Q73b32 bacillus ce
36	699	57.1	235	2 Q81T09	Q81t09 bacillus an
37	699	57.1	235	2 Q6HL92	Q6hl92 bacillus th
38	698	57.0	235	2 Q81FV5	Q81fv5 bacillus ce
39	693.5	56.6	233	1 DEOD_BACSU	Q34925 bacillus su
40	690	56.3	236	1 DEOD_BUCAP	Q8k937 buchnera ap
41	675.5	55.1	235	2 Q8ENY0	Q8eny0 oceanobacil
42	675	55.1	232	2 Q8R973	Q8r973 thermoanaer
43	675	55.1	233	1 DEOD_HELPY	P56463 helicobacte
44	675	55.1	305	2 Q9RSF8	Q9rsf8 deinococcus
45	671.5	54.8	233	2 Q92AF2	Q92af2 listeria in

ALIGNMENTS

RESULT 1
Q7UA16 ID Q7UA16 PRELIMINARY; PRT; 239 AA.
AC Q7UA16;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DE Purine-nucleoside phosphorylase.
GN Name=deod; OrderedLocusNames=S4687;
OS Shigella flexneri.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Shigella.
OX NCBI_TaxID=623;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=2457T;
RX MEDLINE=22590274; PubMed=12704152;
DOI=10.1128/JAI.71.5.2775-2786.2003;
RA Wei J., Goldberg M.B., Burland V., Venkatesan M.M., Deng W.,
Fournier G., Mayhew G.F., Plunkett G. III, Rose D.J., Darling A.,
Mau B., Perna N.T., Payne S.M., Runyen-Janecky L.J., Zhou S.,
Schwartz D.C., Blattner F.R.;
RA "Complete genome sequence and comparative genomics of Shigella
flexneri serotype 2a strain 2457T".
RT Infect. Immun. 71:2775-2786(2003).
DR EMBL; AE016993; AAP19605.1; -.
DR HSP; P09743; 1K9S.
DR GO; GO:0003824; F:catalytic activity; IEA.
DR GO; GO:0004731; F:purine-nucleoside phosphorylase activity; IEA.
DR GO; GO:0009116; F:nucleoside metabolism; IEA.
DR InterPro; IPR004402; Deod.
DR Pfam; PF01048; PNP_UDP_1; 1.
DR TIGRPFAMS; TIGR00107; deod; 1.
DR PROSITE; PS01232; PNP_UDP_1; 1.
SQ SEQUENCE 239 AA; 25936 MW; FA50743A44P87781 CRC64;

Query Match		99.7%;	Score 1221;	DB 2;	Length 239;
Best Local Similarity		99.2%;	Pred. No. 2.3e-92;		
Matches 237;		Conservative 2;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	1	MATPHINAEMGDFADVLMF	Q7UA16	Q7UA16	Q7UA16
Db	1	MATPHINAEMGDFADVLMF	Q7UA16	Q7UA16	Q7UA16
Qy	61	MHGCVGPSCSYIKELITD	Q7UA16	Q7UA16	Q7UA16
Db	61	MHGCVGPSCSYIKELITD	Q7UA16	Q7UA16	Q7UA16
Qy	121	FKQHDFAAIAIDFQVRNA	Q7UA16	Q7UA16	Q7UA16
Db	121	FKQHDFAAIAIDFQVRNA	Q7UA16	Q7UA16	Q7UA16
Qy	181	MEAAIGVGAEEFGAKALT	Q7UA16	Q7UA16	Q7UA16
Db	181	MEAAIGVGAEEFGAKALT	Q7UA16	Q7UA16	Q7UA16

```
|||||
Db 181 MEAAGIYGVAAEFGAKALITCTVSDHIRTAEQTAAERQTTFNMDIKIALESVLLGDKE 239
|||||

RESULT 2
Q83P00
ID Q83P00 PRELIMINARY; PRT; 258 AA.
AC Q83P00;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-MAR-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Purine-nucleoside phosphorylase.
GN Name=deob; OrderedLocusNames=SP4416;
OS Shigella flexneri.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Shigella.
OX NCBI_TaxID=623;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=301 / Serotype 2a;
RX MEDLINE=22272406; PubMed=12384590; DOI=10.1093/nar/gkf566;
RA Jin Q., Yuan Z., Xu J., Wang Y., Shen Y., Lu W., Wang J., Liu H.,
RA Yang J., Yang F., Zhang X., Zhang J., Yang G., Wu H., Qu D., Dong J.,
RA Sun L., Xue Y., Zhao A., Gao Y., Zhu J., Kan B., Ding K., Chen S.,
RA Cheng H., Yao Z., He B., Chen R., Ma D., Qiang B., Wen Y., Hou Y.,
RA Yu J.;
RT "Genome sequence of Shigella flexneri 2a: insights into pathogenicity
RT through comparison with genomes of Escherichia coli K12 and O157."
RL Nucleic Acids Res. 30:4432-4441 (2002).
DR EMBL; AF015447; AAN45831.1; -.
DR HSSP; P09743; 1K9S.
DR GO; GO:0003824; F:catalytic activity; IEA.
DR GO; GO:0004731; F:purine-nucleoside phosphorylase activity; IEA.
DR GO; GO:0009116; P:nucleoside metabolism; IEA.
DR InterPro; IPR004402; Deob.
DR InterPro; IPR000845; PNP_UDP.
DR TIGRFAMs; TIGR00107; deob; 1.
DR PROSITE; PS01232; PNP_UDP_1; 1.
DR Complete proteome.
KW Complete proteome.
SQ SEQUENCE 258 AA; 27956 MW; 3E0BCAA959374365 CRC64;

Query Match 99.7%; Score 1221; DB 2; Length 258;
Best Local Similarity 99.2%; Pred. No. 2.5e-92;
Matches 237; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 MATPHINAEMGDPADVVLMPGDPPLRAKYIAETFLDAREVNNVRGMLGFTGYGKRKISV 60
Db 20 MATPHINAEMGDPADVVLMPGDPPLRAKYIAETFLDAREVNNVRGMLGFTGYGKRKISV 79
|||||

QY 61 MGHGVGIPSCSIYTKELITDFGVKKIIRVSGCAVLPHVKLRDVLVGMGACTDSKNRIR 120
Db 80 MGHGMGIPSCSIYTKELITDFGVKKIIRVSGCAVLPHVKLRDVLVGMGACTDSKNRIR 139
|||||

QY 121 FKDHFAAIAADFDMVNNVDAAKALGIDARVGNLFSADLFYSPDGMFDMVEKYGILGVE 180
Db 140 FKDHFAAIAADFDMVNNVDAAKALGIDARVGNLFSADLFYSPDGMFDMVEKYGILGVE 199
|||||

QY 181 MEAAGIYGVAAEFGAKALITCTVSDHIRTAEQTAAERQTTFNMDIKIALESVLLGDKE 239
Db 200 MEAAGIYGVAAEFGAKALITCTVSDHIRTAEQTAAERQTTFNMDIKIALESVLLGDKE 258
|||||

RESULT 3
Q8FA51
ID Q8FA51 PRELIMINARY; PRT; 258 AA.
AC Q8FA51;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Purine nucleoside phosphorylase (EC 2.4.2.1).
GN Name=deob; OrderedLocusNames=c5468;
OS Escherichia coli O6.
|||||
```

```
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=217992;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=06:H1 / CFT073 / ATCC 700928;
RX MEDLINE=22388234; PubMed=12471157; DOI=10.1073/pnas.252529799;
RA Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,
RA Rasko D., Buckles E.L., Liou S.-R., Boutin A., Hackett J., Stroud D.,
RA Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,
RA Mobley H.L.T., Donnenberg M.S., Blattner F.R.;
RT "Extensive mosaic structure revealed by the complete genome sequence
RT of uropathogenic Escherichia coli."
RL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024 (2002).
DR EMBL; AF016772; AAN83888.1; -.
DR HSSP; P09743; 1K9S.
DR GO; GO:0004731; F:purine-nucleoside phosphorylase activity; IEA.
DR GO; GO:0016757; F:transferase activity, transferring glycosyl. .; IEA.
DR GO; GO:0009116; P:nucleoside metabolism; IEA.
DR InterPro; IPR004402; Deob.
DR InterPro; IPR000845; PNP_UDP.
DR Pfam; PF01048; PNP_UDP_1; 1.
DR TIGRFAMs; TIGR00107; deob; 1.
DR PROSITE; PS01232; PNP_UDP_1; 1.
DR Complete proteome; Glycosyltransferase; Transferase.
KW Complete proteome; Glycosyltransferase; Transferase.
SQ SEQUENCE 258 AA; 27942 MW; EC05C8DFAB3742C1 CRC64;

Query Match 99.7%; Score 1221; DB 2; Length 258;
Best Local Similarity 99.2%; Pred. No. 2.5e-92;
Matches 237; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 MATPHINAEMGDPADVVLMPGDPPLRAKYIAETFLDAREVNNVRGMLGFTGYGKRKISV 60
Db 20 MATPHINAEMGDPADVVLMPGDPPLRAKYIAETFLDAREVNNVRGMLGFTGYGKRKISV 79
|||||

QY 61 MGHGVGIPSCSIYTKELITDFGVKKIIRVSGCAVLPHVKLRDVLVGMGACTDSKNRIR 120
Db 80 MGHGMGIPSCSIYTKELITDFGVKKIIRVSGCAVLPHVKLRDVLVGMGACTDSKNRIR 139
|||||

QY 121 FKDHFAAIAADFDMVNNVDAAKALGIDARVGNLFSADLFYSPDGMFDMVEKYGILGVE 180
Db 140 FKDHFAAIAADFDMVNNVDAAKALGIDARVGNLFSADLFYSPDGMFDMVEKYGILGVE 199
|||||

QY 181 MEAAGIYGVAAEFGAKALITCTVSDHIRTAEQTAAERQTTFNMDIKIALESVLLGDKE 239
Db 200 MEAAGIYGVAAEFGAKALITCTVSDHIRTAEQTAAERQTTFNMDIKIALESVLLGDKE 258
|||||

RESULT 4
DEOD_ECOLI
ID DEOD_ECOLI STANDARD; PRT; 238 AA.
AC P09743;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 25-JAN-2005 (Rel. 46, Last annotation update)
DE Purine nucleoside phosphorylase (EC 2.4.2.1) (Inosine phosphorylase)
DE (PNP).
GN Name=deob; Synonyms=pup; OrderedLocusNames=b4384, z5986, ECs5343;
OS Escherichia coli, and
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562, 83334;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=91334430; PubMed=1714590;
RA Hersfield M.S., Chaffee S., Koro-Johnson L., Mary A., Smith A.A.,
RA Short S.A.;
RT "Use of site-directed mutagenesis to enhance the epitope-shielding
RT effect of covalent modification of proteins with polyethylene
RT glycol."
RL Proc. Natl. Acad. Sci. U.S.A. 88:7185-7189 (1991).
```


FT TURN 173 174
 FT STRAND 177 179
 FT HELIX 182 192
 FT TURN 193 193
 FT STRAND 195 205
 FT TURN 206 208
 FT STRAND 211 211
 FT HELIX 214 219
 FT TURN 223 236
 SQ SEQUENCE 238 AA; C39ACFC1C817AC83E CRC64;
 Query Match 99.3%; Score 1217; DB 1; Length 238;
 Best Local Similarity 99.6%; Pred. No. 4.9e-92;
 Matches 237; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 QY 2 ATPHINAEMGDFADVVLMPGDPPLRAKYIAETPLEDAREVNNVRLGMLGFTGTYGKRKISVM 61
 DB 1 ATPHINAEMGDFADVVLMPGDPPLRAKYIAETPLEDAREVNNVRLGMLGFTGTYGKRKISVM 60
 QY 62 GHGVGIPSCSIYTKELITDFGVKKIIRVSGCGAVLPHVKLRDVLVIGMGACTDSKVNRI 121
 DB 61 GHGVGIPSCSIYTKELITDFGVKKIIRVSGCGAVLPHVKLRDVLVIGMGACTDSKVNRI 120
 QY 122 KDHDFAAIAFDPMVRNAVDAKALGIDARVGNLFSADLFYSPDGMFVMEKYGILGVEM 181
 DB 121 KDHDFAAIAFDPMVRNAVDAKALGIDARVGNLFSADLFYSPDGMFVMEKYGILGVEM 180
 QY 182 EAAGIYGVAAEFGAKALITCTVSDHIRTHTHEQTAAERQTTFNMDIKIALESVLLGDKE 239
 DB 181 EAAGIYGVAAEFGAKALITCTVSDHIRTHTHEQTAAERQTTFNMDIKIALESVLLGDKE 238

RESULT 5
 Q8ZJ02 PRELIMINARY; PRT; 239 AA.
 ID Q8ZJ02
 AC Q8ZJ02; Q7CAT5;
 DT 01-MAR-2002 (TrEMBLrel. 20, Created)
 DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
 DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
 DE Purine nucleoside phosphorylase (EC 2.4.2.1).
 GN Name=deop; Synonyms=pup; OrderedLocusNames=STV4921, t4613;
 OS Salmonella typhi.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Salmonella.
 OX NCBI_TaxID=601;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CT18;
 RX MEDLINE=21534947; PubMed=11677608; DOI=10.1038/35101607;
 RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,
 RA Churcher C.M., Mungall K.L., Bentley S.D., Holden M.T.G., Sebaihia M.,
 RA Baker S., Basham D., Brooks K., Chillingworth T., Connor P.,
 RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
 RA Feltwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels K.,
 RA Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,
 RA Quail M.A., Rutherford K.M., Simmonds M., Skelton J., Stevens K.,
 RA Whitehead S., Barrall B.G.;
 RA "Complete genome sequence of a multiple drug resistant Salmonella
 RT enterica serovar Typhi CT18";
 RL Nature 413:848-852(2001).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Ty2 / ATCC 700931;
 RX MEDLINE=22531367; PubMed=12644504;
 RX DOI=10.1128/JB.185.7.2330-2337.2003;
 RA Deng W., Liou S.-R., Plunkett G. III, Mayhew G.F., Rose D.J.,
 RA Burland V., Kodoyianni V., Schwartz D.C., Blattner F.R.;
 RT "Comparative genomics of Salmonella enterica serovar Typhi strains Ty2
 and CT18";
 RL J. Bacteriol. 185:2330-2337(2003).
 DR EMBL; AL627284; CAD03405.1; -.
 DR EMBL; AE016849; AA072045.1; -.
 DR HSSP; P09743; 1PR6.

DR GO; GO:0004731; F:purine-nucleoside phosphorylase activity; IEA.
 DR GO; GO:0016757; P:transferase activity, transferring glycosyl. . .; IEA.
 DR GO; GO:0009116; P:nucleoside metabolism; IEA.
 DR PFam; PF01048; PNP_UDP_1; 1.
 DR TIGRFAMs; TIGR00107; deod; 1.
 DR PROSITE; PS01232; PNP_UDP_1; 1.
 KW Complete proteome; Glycosyltransferase; Transferase.
 SQ SEQUENCE 239 AA; 25979 MW; 7A4AFD72C646859E CRC64;
 Query Match 96.9%; Score 1187; DB 2; Length 239;
 Best Local Similarity 96.7%; Pred. No. 1.4e-89;
 Matches 231; Conservative 4; Mismatches 4; Indels 0; Gaps 0;
 QY 1 MATPHINAEMGDFADVVLMPGDPPLRAKYIAETPLEDAREVNNVRLGMLGFTGTYGKRKISV 60
 DB 1 MATPHINAEMGDFADVVLMPGDPPLRAKYIAETPLEDAREVNNVRLGMLGFTGTYGKRKISV 60
 QY 61 MHGVGIPSCSIYTKELITDFGVKKIIRVSGCGAVLPHVKLRDVLVIGMGACTDSKVNRI 120
 DB 61 MHGVGIPSCSIYTKELITDFGVKKIIRVSGCGAVLPHVKLRDVLVIGMGACTDSKVNRI 120
 QY 121 FKDHFAAIAFDPMVRNAVDAKALGIDARVGNLFSADLFYSPDGMFVMEKYGILGV 180
 DB 121 FKDHFAAIAFDPMVRNAVDAKALGIDARVGNLFSADLFYSPDGMFVMEKYGILGV 180
 QY 181 MEAAGIYGVAAEFGAKALITCTVSDHIRTHTHEQTAAERQTTFNMDIKIALESVLLGDKE 239
 DB 181 MEAAGIYGVAAEFGAKALITCTVSDHIRTHTHEQTAAERQTTFNMDIKIALESVLLGDKE 239
 RESULT 6
 Q8ZJV7 PRELIMINARY; PRT; 239 AA.
 ID Q8ZJV7
 AC Q8ZJV7;
 DT 01-MAR-2002 (TrEMBLrel. 20, Created)
 DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
 DE Purine nucleoside phosphorylase (EC 2.4.2.1).
 GN Name=deop; OrderedLocusNames=STW4570;
 OS Salmonella typhimurium.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Salmonella.
 OX NCBI_TaxID=602;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=LT2;
 RX MEDLINE=21534948; PubMed=11677609; DOI=10.1038/35101614;
 RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
 RA Courtney L., Porwollik S., Ali J., Dante M., Du P., Hou S., Layman D.,
 RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
 RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
 RA Waterston R., Wilson R.K.;
 RA "Complete genome sequence of Salmonella enterica serovar Typhimurium
 RT LT2";
 RL Nature 413:852-856(2001).
 RL EMBL; AE008915; AAL23385.1; -.
 DR HSSP; P09743; 1PR6.
 DR GO; GO:0004731; F:purine-nucleoside phosphorylase activity; IEA.
 DR GO; GO:0016757; P:transferase activity, transferring glycosyl. . .; IEA.
 DR GO; GO:0009116; P:nucleoside metabolism; IEA.
 DR PFam; PF01048; PNP_UDP_1; 1.
 DR TIGRFAMs; TIGR00107; deod; 1.
 DR PROSITE; PS01232; PNP_UDP_1; 1.
 KW Complete proteome; Glycosyltransferase; Transferase.
 SQ SEQUENCE 239 AA; 25978 MW; 90ED901A61814796 CRC84;
 Query Match 96.5%; Score 1182; DB 2; Length 239;
 Best Local Similarity 96.2%; Pred. No. 3.7e-89;
 Matches 230; Conservative 5; Mismatches 4; Indels 0; Gaps 0;
 QY 1 MATPHINAEMGDFADVVLMPGDPPLRAKYIAETPLEDAREVNNVRLGMLGFTGTYGKRKISV 60
 DB 1 MATPHINAEMGDFADVVLMPGDPPLRAKYIAETPLEDAREVNNVRLGMLGFTGTYGKRKISV 60

```

Qy 61 MGHGVGIPSCSIYTKELITDFGVKKIIRVSGCAGVLPVHVKLRDVVVGACTDSKVNIR 120
D 61 MGHGVGIPSCSIYTKELITDFGVKKIIRVSGCAGVLPVHVKLRDVVVGACTDSKVNIR 120
Qy 121 FKDHDFAAIADPDMVRNVAADAAKALGIDARVGNLFSADLFYSPDGMFDMVEKYGILGVE 180
D 121 FKDHDFAAIADPDMVRNVAADAAKALGIDARVGNLFSADLFYSPDGMFDMVEKYGILGVE 180
Qy 181 MEAGIYGAEEFGAKALITCTVSDHIRTHTQTTAAERQTTFNNDMKIALSVLLGD 237
D 181 MEAGIYGAEEFGAKALITCTVSDHIRTHTQTTAAERQTTFNNDMKIALSVLLGD 237

RESULT 7
ID Q66EV7 PRELIMINARY; PRT; 239 AA.
AC Q66EV7;
DT 25-OCT-2004 (TrEMBLrel. 28, Created)
DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DE Purine-nucleoside phosphorylase (EC 2.4.2.1)
GN Name=deod; Synonyms=pup; ORFNames=YPTB0584;
OS Yersinia pseudotuberculosis IP 32953.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Yersinia.
OX NCBI_TaxID=273123;
RN [1]
RP SEQUENCE FROM N.A.
RX PubMed=15358858;
RA Chain P.S.G., Carniel E., Larimer F.W., Lamerdin J., Stoutland P.O.,
RA Regala W.M., Georgescu A.M., Verges L.M., Land M.L., Motin L.V.,
RA Brubaker R.R., Fowler J., Hinnebusch B.J., Marceau M., Medigue C.,
RA Simonet M., Chenal-Francisque V., Souza B., Dacheux D., Elliott J.M.,
RA Derbise A., Hauser L.J., Garcia E.;
RA "Insights into the genome evolution of Yersinia pestis through whole
RT genome comparison with Yersinia pseudotuberculosis."
RL Proc. Natl. Acad. Sci. U.S.A. 101:13826-13831(2004).
DR EMBL; BX936398; CAH19824.1; -.
DR GO; GO:0016757; F:transferase activity, transferring glycosyl. .; IEA.
DR InterPro; IPR004402; Deod.
DR InterPro; IPR000845; PNP_UDP.
DR Pfam; PF01048; PNP_UDP_1; 1.
DR TIGRFAMs; TIGR00107; deod; 1.
DR PROSITE; PS01232; PNP_UDP_1; 1.
KW Glycosyltransferase; Transferase.
SQ SEQUENCE 239 AA; 25936 MW; 344A15BPAA7P7975 CRC64;

Query Match 91.0%; Score 1115; DB 2; Length 239;
Best Local Similarity 91.1%; Pred. No. 1.2e-83;
Matches 216; Conservative 11; Mismatches 10; Indels 0; Gaps 0;

Qy 1 MATHINAEAGDFADVLMPCGDLRAKYIAETFLDAREVNVNVRGMLGFTGYGRKISV 60
D 1 MATHINAEAGDFADVLMPCGDLRAKYIAETFLDAREVNVNVRGMLGFTGYGRKISV 60
Qy 61 MGHGVGIPSCSIYTKELITDFGVKKIIRVSGCAGVLPVHVKLRDVVVGACTDSKVNIR 120
D 61 MGHGVGIPSCSIYTKELITDFGVKKIIRVSGCAGVLPVHVKLRDVVVGACTDSKVNIR 120
Qy 121 FKDHDFAAIADPDMVRNVAADAAKALGIDARVGNLFSADLFYSPDGMFDMVEKYGILGVE 180
D 121 FKDHDFAAIADPDMVRNVAADAAKALGIDARVGNLFSADLFYSPDGMFDMVEKYGILGVE 180
Qy 181 MEAGIYGAEEFGAKALITCTVSDHIRTHTQTTAAERQTTFNNDMKIALSVLLGD 237
D 181 MEAGIYGAEEFGAKALITCTVSDHIRTHTQTTAAERQTTFNNDMKIALSVLLGD 237

RESULT 8
Q82IQ2 PRELIMINARY; PRT; 239 AA.
ID Q82IQ2

```

```

Q82IQ2;
Qy 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Purine nucleoside phosphorylase (EC 2.4.2.1)
GN Name=deod; Synonyms=b4384, pup; OrderedLocusNames=YPO0440;
OS Yersinia pestis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Yersinia.
OX NCBI_TaxID=632;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=CO-92 / Biovar Orientalis;
RX MEDLINE=21470413; PubMed=11586360; DOI=10.1038/35097083;
RA Parkhill J., Wren B.W., Thomson N.R., Titball R.W., Holden M.T.G.,
RA Prentice M.B., Sebahia M., James K.D., Churcher C.M., Mungall K.L.,
RA Baker S., Basham D., Bentley S.D., Brooks K., Cerdeno-Tarraga A.-M.,
RA Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,
RA Feltwell T., Hamlin N., Holroyd S., Jagsels K., Karlyshev A.V.,
RA Leather S., Moule S., Oyston P.C.F., Quail M.A., Rutherford K.M.,
RA Simmonds M., Skelton J., Stevens K., Whitehead S., Barrell B.G.;
RT "Genome sequence of Yersinia pestis, the causative agent of plague.";
RL Nature 413:523-527(2001).
DR EMBL; AJ414142; CAC89297.1; -.
DR PIR; AF0054; AF0054.
DR HSP; P09743; IPR6.
DR GO; GO:0004731; F:purine-nucleoside phosphorylase activity; IEA.
DR GO; GO:0016757; F:transferase activity, transferring glycosyl. .; IEA.
DR GO; GO:0009116; P:nucleoside metabolism; IEA.
DR Pfam; PF01048; PNP_UDP_1; 1.
DR TIGRFAMs; TIGR00107; deod; 1.
DR PROSITE; PS01232; PNP_UDP_1; 1.
KW Complete proteome; Glycosyltransferase; Transferase.
SQ SEQUENCE 239 AA; 25876 MW; 684A15BFBBA9A209 CRC64;

Query Match 90.3%; Score 1106; DB 2; Length 239;
Best Local Similarity 90.7%; Pred. No. 6.5e-83;
Matches 215; Conservative 11; Mismatches 11; Indels 0; Gaps 0;

Qy 1 MATHINAEAGDFADVLMPCGDLRAKYIAETFLDAREVNVNVRGMLGFTGYGRKISV 60
D 1 MATHINAEAGDFADVLMPCGDLRAKYIAETFLDAREVNVNVRGMLGFTGYGRKISV 60
Qy 61 MGHGVGIPSCSIYTKELITDFGVKKIIRVSGCAGVLPVHVKLRDVVVGACTDSKVNIR 120
D 61 MGHGVGIPSCSIYTKELITDFGVKKIIRVSGCAGVLPVHVKLRDVVVGACTDSKVNIR 120
Qy 121 FKDHDFAAIADPDMVRNVAADAAKALGIDARVGNLFSADLFYSPDGMFDMVEKYGILGVE 180
D 121 FKDHDFAAIADPDMVRNVAADAAKALGIDARVGNLFSADLFYSPDGMFDMVEKYGILGVE 180
Qy 181 MEAGIYGAEEFGAKALITCTVSDHIRTHTQTTAAERQTTFNNDMKIALSVLLGD 237
D 181 MEAGIYGAEEFGAKALITCTVSDHIRTHTQTTAAERQTTFNNDMKIALSVLLGD 237

RESULT 9
Q8CZ9 PRELIMINARY; PRT; 241 AA.
ID Q8CZ9 Q74PY9;
AC Q8CZ9;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DE Purine nucleoside phosphorylase.
GN Name=deod; OrderedLocusNames=YP3742, Y3740;
OS Yersinia pestis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Yersinia.
OX NCBI_TaxID=632;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=KIM5 / Biovar Mediaevalis;
RX MEDLINE=22137863; PubMed=12142430;

```

DOI=10.1128/JB.184.16.4601-4611.2002;
RA Deng W., Burland V., Plunkett G. III, Boutin A., Mayhew G.F., Liss P.,
RA Perna N.T., Rose D.J., Mau B., Zhou S., Schwartz D.C.,
RA Fetherston J.D., Lindler L.E., Brubaker R.R., Plano G.V.,
RA Straley S.C., McDonough K.A., Nilles M.D., Matson J.S., Blattner F.R.,
RA Perry R.D.;
RT "Genome sequence of *Yersinia pestis* KIM.";
RN J. Bacteriol. 184:4601-4611(2002).
RP SEQUENCE FROM N.A.
RC STRAIN=91001 / Biovar Mediaevalis;
RA Song Y., Tong Z., Wang L., Han Y., Zhang J., Pei D., Wang J., Zhou D.,
RA Han Y., Pang X., Zhai J., Chen F., Qin H., Wang J., Li S., Guo Z.,
RA Ye C., Du Z., Lin W., Wang J., Yu J., Yang H., Wang J., Huang P.,
RA Yang R.;
RL Submitted (APR-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE013977; AAM87286.1; -;
DR EMBL; AE017141; AAS63890.1; -;
DR HSSP; P09743; 1PR6.
DR GO; GO:0003824; F:catalytic activity; IEA.
DR GO; GO:0004731; F:purine-nucleoside phosphorylase activity; IEA.
DR GO; GO:0009116; P:nucleoside metabolism; IEA.
DR InterPro; IPR004402; Deod.
DR Pfam; PF01048; PNP_UDP_1; 1.
DR TIGRFAMs; TIGR00107; deod; 1.
DR PROSITE; PS01232; PNP_UDP_1; 1.
DR TIGR00107; deod; 1.
DR PROSITE; PS01232; PNP_UDP_1; 1.
SQ SEQUENCE 241 AA; 26120 MW; 1.797D09F97072CFA9 CRC64;

Query Match 90.3%; Score 1106; DB 2; Length 241;
Best Local Similarity 90.7%; Pred. No. 6.6e-83;
Matches 215; Conservative 11; Mismatches 11; Indels 0; Gaps 0;

Qy 1 MATPHINAEMGDFADVVLMPCGDPRAKYIAETFLDAREVNNVGMGLFTGTGKRSIV 60
Db 3 MATPHINAEMGDFADVVLMPCGDPRAKYIAETFLDAREVNNVGMGLFTGTGKRSIV 62

Qy 61 MGHGVGIPSCSIYTKELITDFGVKKIIRVSCGAVLPHVKLRDVGWGMGACTDSKNRIR 120
Db 63 MGHGMGIPSCSIYAKELITDFGVKKIIRVSCGAVRDKLRDVGWGMGACTDSKNRMR 122

Qy 121 FKDHFAAIAADFDMVRNAVDAAKALGIDARVGNLFSADLPSPDGMFDMVEKYGILGVE 180
Db 123 FKDHVAAIAADFEMTRNAVDAAKAKGVNRVGNLFSADLPFTDPQMFDMVEKYGILGVE 182

Qy 181 MEAGIYGVAAEFCAKALITCTVSDHRTTAAEROTTENDMIKIALESVLLGD 237
Db 183 MEAGICGVAAEFCAKALITCTVSDHRTTAAEROTTENDMIKIALESVLLGD 239

RESULT 10
Q6D989 PRELIMINARY; PRT; 239 AA.
AC Q6D989
DT 25-OCT-2004 (TrEMBLrel. 28, Created)
DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE Purine nucleoside phosphorylase (EC 2.4.2.1).
GN Name=deob; Synonym=pup; OrderedLocustNames=ECA0730;
OS *Erwinia carotovora* (subsp. atroseptica) (Pectobacterium atrosepticum).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Pectobacterium.
OX NCBI_TaxID=29471;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SCRI 1043 / ATCC BAA-672;
RX PubMed=15263089; DOI=10.1073/pnas.0402424101;
RA Bell K.S., Sebahia M., Pritchard L., Holden M.T.G., Hyman L.J.,
RA Holvea M.C., Thomson N.R., Bentley S.D., Churcher L.J.C., Mungall K.,
RA Atkin R., Bason N., Brooks K., Chillingworth T., Clark K., Doggett J.,
RA Fraser A., Hance Z., Hauser H., Jagels K., Moule S., Norbertczak H.,
RA Ormond D., Price C., Quail M.A., Sanders M., Walker D., Whitehead S.,
RA Salmund G.P.C., Birch P.R.J., Parkhill J., Toth I.K.;

"Genome sequence of the enterobacterial phytopathogen *Erwinia carotovora* subsp. atroseptica and characterization of virulence factors.";
RL Proc. Natl. Acad. Sci. U.S.A. 101:11105-11110(2004).
DR EMBL; BX950851; CAG73644.1; -;
DR GO; GO:0004731; F:purine-nucleoside phosphorylase activity; IEA.
DR GO; GO:0016757; F:transferase activity, transferring glycosyl. .; IEA.
DR GO; GO:0009116; P:nucleoside metabolism; IEA.
DR InterPro; IPR004402; Deod.
DR InterPro; IPR000845; PNP_UDP.
DR Pfam; PF01048; PNP_UDP_1; 1.
DR TIGRFAMs; TIGR00107; deod; 1.
DR PROSITE; PS01232; PNP_UDP_1; 1.
KW Complete proteome; Glycosyltransferase; Transferase.
SQ SEQUENCE 239 AA; 26038 MW; AC910A030E86170A CRC64;

Query Match 89.4%; Score 1095; DB 2; Length 239;
Best Local Similarity 88.7%; Pred. No. 5.2e-82;
Matches 212; Conservative 13; Mismatches 14; Indels 0; Gaps 0;

Qy 1 MATPHINAEMGDFADVVLMPCGDPRAKYIAETFLDAREVNNVGMGLFTGTGKRSIV 60
Db 1 MATPHINAEMGDFADVVLMPCGDPRAKYIAETFLDAREVNNVGMGLFTGTGKRSIV 60

Qy 61 MGHGVGIPSCSIYTKELITDFGVKKIIRVSCGAVLPHVKLRDVGWGMGACTDSKNRIR 120
Db 61 MGHGMGIPSCSIYAKELITDFGVKKIIRVSCGAVRDKLRDVGWGMGACTDSKNRMR 120

Qy 121 FKDHFAAIAADFDMVRNAVDAAKALGIDARVGNLFSADLPSPDGMFDMVEKYGILGVE 180
Db 121 FKDHVAAIAADFDMVRNAVDAAKARDVSRVGNLFSADLPFTDPQMFDMVEKYGILGVE 180

Qy 181 MEAGIYGVAAEFCAKALITCTVSDHRTTAAEROTTENDMIKIALESVLLGDKE 239
Db 181 MEAGIYGVAAEFCAKALITCTVSDHRTTAAEROTTENDMIKIALESVLLGDNE 239

RESULT 11
Q8KRT5 PRELIMINARY; PRT; 238 AA.
AC Q8KRT5
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE Purine nucleoside phosphorylase.
GN Name=deob;
OS *Xenorhabdus nematophilus*.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Xenorhabdus.
OX NCBI_TaxID=628;
RN [1]
RP SEQUENCE FROM N.A.
RA He H., Forst S.;
RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 19061;
RA Orchard S.S., Goodrich-Blair H.;
RL Submitted (FEB-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF525420; AAM91930.1; -;
DR EMBL; AY556400; AAT40584.1; -;
DR HSSP; P09743; 1PKE.
DR GO; GO:0003824; F:catalytic activity; IEA.
DR GO; GO:0004731; F:purine-nucleoside phosphorylase activity; IEA.
DR GO; GO:0009116; P:nucleoside metabolism; IEA.
DR InterPro; IPR004402; Deod.
DR InterPro; IPR000845; PNP_UDP.
DR Pfam; PF01048; PNP_UDP_1; 1.
DR TIGRFAMs; TIGR00107; deod; 1.
DR PROSITE; PS01232; PNP_UDP_1; 1.
SQ SEQUENCE 238 AA; 26006 MW; 682D984DE51441CA CRC64;

Query Match 88.7%; Score 1086; DB 2; Length 238;

```
Best Local Similarity 88.2%; Pred. No. 2.9e-81;
Matches 209; Conservative 15; Mismatches 13; Indels 0; Gaps 0;

Qy 1 MATPHINAEMGDFADVLMPCGDPPLRAKYIAETFLDAREVNNVRGMLGFTGTYGKRKISV 60
Db 1 MATPHINAEMGDFADVLMPCGDPPLRAQYIAETFLDARQVNNVRGMLGFTGTYGKRISV 60

Qy 61 MGHGVGIPSCSIYTKELITDFGVKKIIRVSGCGAVLPHVKLRDVRVVGAGCTDSKVNRI 120
Db 61 MGHGMGIPSCSIYAKELITDFGVKKIIRIGSCGAVSDVKIRDVVVGAGCTDSKVNMR 120

Qy 121 FKDHDFAAIADPDMVRNADAAKALGIDARVGNLFSADLFYSPDGENFDMVKYGLGVE 180
Db 121 FKDHDFAAIADPDMVRNADAAKALGIDARVGNLFSADLFYSPDGENFDMVKYGLGVE 180

Qy 121 FKDHDFAAIADPDMVRNADAAKALGIDARVGNLFSADLFYSPDGENFDMVKYGLGVE 180
Db 121 FKDHDFAAIADPDMVRNADAAKALGIDARVGNLFSADLFYSPDGENFDMVKYGLGVE 180

Qy 181 MEAGIYGVAAEFCAKALITCTVSDHRTHTQTTAAERQTTFNDMIKIALESVLLGD 237
Db 181 MEAGIYGVAAEFCAKALITCTVSDHRTHTQTTAAERQTTFNDMIKIALESVLLGD 237

RESULT 12
DEOD KLEPN STANDARD; PRT; 239 AA.
AC Q59482;
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Purine nucleoside phosphorylase (EC 2.4.2.1) (Inosine phosphorylase)
DE (PNP)
GN Name=deod; Synonyms=ppp;
OS Klebsiella pneumoniae;
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Klebsiella.
OX NCBI_TaxID=573;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=LFI202;
RA MEDLINE=96069952; PubMed=8534998;
RA Takehana M., Ling F., Izawa S., Inoue Y., Kimura A.;
RT "Molecular cloning and nucleotide sequence of purine nucleoside
RT phosphorylase and uridine phosphorylase genes from Klebsiella sp.";
RL Biosci. Biotechnol. Biochem. 59:1987-1990(1995).
CC -1- FUNCTION: Cleavage of guanosine or inosine to respective bases and
CC sugar-1-phosphate molecules.
CC -1- CATALYTIC ACTIVITY: Purine nucleoside + phosphate = purine +
CC alpha-D-ribose 1-phosphate.
CC -1- SUBUNIT: Homohexamer (By similarity).
CC -1- SIMILARITY: Belongs to the PNP/UDP phosphorylase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.ebi.ac.uk/announcements/
CC or send an email to license@ebi.ac.uk).
CC
CC EMBL; X87881; CAA61136.1; --
CC HSP; P09743; 1K9S
CC InterPro; IPR004402; Deod.
CC InterPro; IPR000845; PNP_UDP.
CC Pfam; PF01048; PNP_UDP_1; 1.
CC TIGRFAMs; TIGR00107; deod; 1.
CC TIGRFAMs; TIGR01718; Uridine-phosphatase; 1.
CC PROSITE; PS01232; PNP_UDP_1; 1.
CC Glycosyltransferase; Transferase.
CC SEQUENCE 239 AA; 26196 MW; BC7514EDF5B27F5 CRC64;

Qy 1 MATPHINAEMGDFADVLMPCGDPPLRAKYIAETFLDAREVNNVRGMLGFTGTYGKRKISV 60
Db 1 MATPHINAEMGDFADVLMPCGDPPLRAKYIAETFLDAREVNNVRGMLGFTGTYGKRKISV 60

Qy 61 MGHGVGIPSCSIYTKELITDFGVKKIIRVSGCGAVLPHVKLRDVRVVGAGCTDSKVNRI 120
Db 61 MGHGMGIPSCSIYAKELITDFGVKKIIRVSGCGAVLPHVKLRDVRVVGAGCTDSKVNRI 120

Qy 121 FKDHDFAAIADPDMVRNADAAKALGIDARVGNLFSADLFYSPDGENFDMVKYGLGVE 180
Db 121 FKDHDFAAIADPDMVRNADAAKALGIDARVGNLFSADLFYSPDGENFDMVKYGLGVE 180

Qy 181 MEAGIYGVAAEFCAKALITCTVSDHRTHTQTTAAERQTTFNDMIKIALESVLLGD 237
Db 181 MEAGIYGVAAEFCAKALITCTVSDHRTHTQTTAAERQTTFNDMIKIALESVLLGD 237

Query Match 88.4%; Score 1085; DB 1; Length 239;
Best Local Similarity 90.6%; Pred. No. 3.5e-81;
Matches 217; Conservative 7; Mismatches 14; Indels 2; Gaps 2;

Qy 1 MATPHINAEMGDFADVLMPCGDPPLRAKYIAETFLDAREVNNVRGMLGFTGTYGKRKISV 60
Db 1 MATPHINAEMGDFADVLMPCGDPPLRAKYIAETFLDAREVNNVRGMLGFTGTYGKRKISV 60

Qy 61 MGHGVGIPSCSIYTKELITDFGVKKIIRVSGCGAVLPHVKLRDVRVVGAGCTDSKVNRI 120
Db 61 MGHGMGIPSCSIYAKELITDFGVKKIIRVSGCGAVLPHVKLRDVRVVGAGCTDSKVNRI 120

Qy 121 FKDHDFAAIADPDMVRNADAAKALGIDARVGNLFSADLFYSPDGENFDMVKYGLGVE 180
Db 121 FKDHDFAAIADPDMVRNADAAKALGIDARVGNLFSADLFYSPDGENFDMVKYGLGVE 180

Qy 181 MEAGIYGVAAEFCAKALITCTVSDHRTHTQTTAAERQTTFNDMIKIALESVLLGD 237
Db 181 MEAGIYGVAAEFCAKALITCTVSDHRTHTQTTAAERQTTFNDMIKIALESVLLGD 237
```

```
Db 1 MATPHINAEMGDFADVLMPCGDPPLRAKYIAETFLDAREVNNVRGMLGFTGTYGKRKISV 60
Qy 60 VMGHGVGIPSCSIYTKELITDFGVKKIIRVSGCGAVLPHVKLRDVRVVGAGCTDSKVNRI 119
Db 61 VMGHGMGIPSCSIYTKELITDFGVKKIIRVSGCGAVLPHVKLRDVRVVGAGCTDSKVNRI 120

Qy 120 RFKDHDFAAIADPDMVRNADAAKALGIDARVGNLFSADLFYSPDGENFDMVKYGLGVE 179
Db 121 RFKDHDFAAIADPDMVRNADAAKALGIDARVGNLFSADLFYSPDGENFDMVKYGLGVE 179

Qy 180 EMEAGIYGVAAEFCAKALITCTVSDHRTHTQTTAAERQTTFNDMIKIALESVLLGD 239
Db 180 EMEAGIYGVAAEFCAKALITCTVSDHRTHTQTTAAERQTTFNDMIKIALESVLLGD 239

RESULT 13
Q7N930 PRELIMINARY; PRT; 238 AA.
AC Q7N930;
DT 01-MAR-2004 (TReMBLrel. 26, Created)
DT 01-MAR-2004 (TReMBLrel. 26, Last sequence update)
DT 01-MAR-2004 (TReMBLrel. 26, Last annotation update)
DE Purine nucleoside phosphorylase (Inosine phosphorylase) (PNP).
DE Name=deod; OrderedLocNames=plu0522;
OS Phototrichum luminescens (subsp. laumondii).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Phototrichum.
OX NCBI_TaxID=141679;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TT01;
RX MEDLINE=22957627; PubMed=14528314;
RA Duchaud E., Rusniok C., Frangeul L., Buchrieser C., Givaudan A.,
RA Taourit S., Bocs S., Bouraux-Eude C., Chandler M., Charles J.-F.,
RA Dassa E., Derose R., Derzelle S., Freysinet G., Gaudriault S.,
RA Medigue C., Lanois A., Powell K., Siguer P., Vincent R., Wingate V.,
RA Zouine M., Glaser P., Boemare N., Danchin A., Kunst F.;
RT "The genome sequence of the entomopathogenic bacterium Phototrichum
RT luminescens."
RL Nat. Biotechnol. 21:1307-1313(2003).
RL EMBL; BX571860; CAE12817.1; --.
DR HSP; P09743; 1PR1.
DR Photolact; plu0522; --.
DR GO; GO:0003824; F:catalytic activity; IEA.
DR GO; GO:0004731; F:purine-nucleoside phosphorylase activity; IEA.
DR GO; GO:0009116; P:nucleoside metabolism; IEA.
DR InterPro; IPR004402; Deod.
DR InterPro; IPR000845; PNP_UDP.
DR Pfam; PF01048; PNP_UDP_1; 1.
DR TIGRFAMs; TIGR00107; deod; 1.
DR PROSITE; PS01232; PNP_UDP_1; 1.
KW Complete proteome.
SQ SEQUENCE 238 AA; 26114 MW; 13651F7C82C49E9A CRC64;

Query Match 87.1%; Score 1067; DB 2; Length 238;
Best Local Similarity 85.7%; Pred. No. 1e-79;
Matches 204; Conservative 18; Mismatches 16; Indels 0; Gaps 0;

Qy 1 MATPHINAEMGDFADVLMPCGDPPLRAKYIAETFLDAREVNNVRGMLGFTGTYGKRKISV 60
Db 1 MATPHINAEMGDFADVLMPCGDPPLRAKYIAETFLDAREVNNVRGMLGFTGTYGKRKISV 60

Qy 61 MGHGVGIPSCSIYTKELITDFGVKKIIRVSGCGAVLPHVKLRDVRVVGAGCTDSKVNRI 120
Db 61 MGHGMGIPSCSIYAKELITDFGVKKIIRVSGCGAVLPHVKLRDVRVVGAGCTDSKVNRI 120

Qy 121 FKDHDFAAIADPDMVRNADAAKALGIDARVGNLFSADLFYSPDGENFDMVKYGLGVE 180
Db 121 FKDHDFAAIADPDMVRNADAAKALGIDARVGNLFSADLFYSPDGENFDMVKYGLGVE 180

Qy 181 MEAGIYGVAAEFCAKALITCTVSDHRTHTQTTAAERQTTFNDMIKIALESVLLGD 238
Db 181 MEAGIYGVAAEFCAKALITCTVSDHRTHTQTTAAERQTTFNDMIKIALESVLLGD 238
```

